

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



THE NEW WORLD OF MENTAL HEALTH

Turning global awareness into global action **PAGES 20 & 25**

GET READING

BOOK REVIEWS SUPPLEMENT

*A farewell to sex, the cosmic
songbook and more*

PAGE 20

BIOGEOCHEMISTRY

GROUNDS FOR OPTIMISM

*A role for smart soils in
global warming mitigation*

PAGE 49

ASTROPHYSICS

NOISY NEIGHBOURS

*Nearby supernovae that
left their mark on Earth*

PAGES 40, 60 & 73

NATURE.COM/NATURE

7 April 2016 £10

Vol. 532, No. 7597



THIS WEEK

EDITORIALS

GLOBAL HEALTH Tackling mental illness needs tangible plans — and cash **p.6**

WORLD VIEW Enmesh social science with research from the start **p.7**



INFECTIOUS DISEASE Viruses morph rapidly inside their mosquito hosts **p.8**

Viral complacency

The first outbreak of yellow fever in Angola in almost 30 years illustrates the danger of a short attention span when confronting epidemic threats.

The World Health Organization (WHO) last week declared that the Ebola epidemic in West Africa no longer represents an international public-health emergency. But as experts also warned last week, there must be no let-up in improving readiness for the next Ebola outbreak — including the nightmare prospect of an epidemic in the megacities of Africa.

An expert panel of researchers convened by the London-based Wellcome Trust and the Center for Infectious Disease Research and Policy at the University of Minnesota, Minneapolis, warned in particular that support for Ebola vaccine research must not be allowed to slip. Although a successful experimental Ebola vaccine has been developed, much work remains to be done if safe and effective vaccines are to be ready in sufficient amounts to quickly nip future outbreaks in the bud.

“As Ebola infection rates come under control it’s a huge concern that complacency sets in, attention moves to more immediate threats, and Ebola vaccine development is left half-finished,” cautioned Jeremy Farrar, director of the Wellcome. Too often in the past, the world has stumbled from epidemic to epidemic, failing to learn the lessons of the last. Emergency responses to the latest threat capture headlines, research and political attention. But too often this attention quickly fades.

Yellow fever, a virus that kills many of those it infects, is just one example of the failure to sustain control efforts. The virus is spread by *Aedes aegypti*, a mosquito adapted to live in urban areas. An ongoing yellow-fever outbreak in Angola, the first for nearly 30 years, began in December and has since spread within the country, infecting at least 490 people and killing 198. It now threatens the wider region.

Yet mass vaccination and intensive mosquito-control programmes largely eliminated this vaccine-preventable mosquito-borne disease by the 1940s. In South America, where the disease was widespread, the mosquito vector was virtually wiped off the map by the 1970s. But the success of mosquito control led to complacency and scale-back.

As a result, *A. aegypti* is now present across more of the continent than before control began. It is similarly resurgent in tropical and subtropical regions worldwide, resulting in sporadic outbreaks of yellow fever in at-risk countries. Moreover, the mosquito’s comeback has fuelled large urban outbreaks of dengue, chikungunya and now Zika viruses, with at least tens of millions of people infected.

There’s a long list of other *Aedes*-borne viruses that are currently restricted to animal reservoirs in the wild. But some of these, including potentially deadly ones, will inevitably establish themselves in cities with *Aedes* mosquitoes. Given rampant urbanization throughout the tropics and subtropics, dense human and *Aedes* populations are ticking time bombs. The failure to sustain *Aedes* control illustrates the need for long-term persistence to curb epidemic threats.

The Angola outbreak has already depleted an international emergency stockpile of 6 million doses of vaccine, leaving authorities scrambling to obtain extra vaccine from national immunization programmes. The WHO and other international agencies launched the

Yellow Fever Initiative in 2006 to reboot yellow-fever mass-vaccination programmes and routine vaccination in the highest-risk African countries. But vaccine stocks are still insufficient, and vaccine coverage in many African countries too low, leaving many vulnerable. The problem, says one official associated with the initiative, is that yellow fever is a “forgotten disease”, which makes it difficult to attract sustained political interest and funding.

“Too often in the past, the world has stumbled from epidemic to epidemic, failing to learn the lessons of the last.”

Following the Ebola epidemic in West Africa, a slew of commissions and reports laid out a broad consensus on what needs to be done for a more proactive and sustained preparedness against epidemic threats. These measures include reinforcing public-health systems, surveillance and diagnostic capacities, and training health workers to identify and respond early to disease outbreaks. Weaknesses in these areas have been

identified as factors that allowed what was a small Ebola outbreak to spiral out of control.

Monitoring viruses in the wild, and a better understanding of how factors such as deforestation, and the hunting and consumption of bushmeat, influence spillover of animal viruses into humans, is key. So, too, is the pre-emptive development of drugs and vaccines against known potential epidemic threats.

The Ebola epidemic has prompted vigorous discussion on all these points, on what shape or form any new global initiatives should take, and where the required multibillion-dollar investment will come from. The big risk is that as the Ebola epidemic fades from memory, the sustained political commitment and funding required will not materialize, and business as usual will resume. That must not be allowed to happen. ■

Safety in neutrons

To boost nuclear security, research reactors must eliminate highly enriched uranium.

Working with the United States, Japan has removed all of the highly enriched uranium (HEU) and the separated plutonium from one of its nuclear reactor facilities, to minimize the risk of theft and use by terrorists. The two countries have now pledged to convert a second research reactor to use safer, low-enriched uranium. These are among the latest in a series of accomplishments that have stemmed from US President Barack Obama’s biennial Nuclear

Security Summit, which wrapped up on 1 April. More than 50 countries attended, most represented by heads of state, making a variety of commitments to reduce the risk of nuclear terrorism.

These projects are also a reminder of just how slow progress has been — and how much remains to be done.

Obama launched the agenda in a 2009 speech in Prague, calling on governments to secure or eliminate all vulnerable nuclear materials in four years. His speech underscored the fact that the threat of a nuclear attack has increased even as the danger of apocalyptic nuclear warfare has receded. Recent revelations that the Islamist terrorist group ISIS may have been targeting a nuclear facility in Belgium make this all too clear.

The initial focus has been on HEU, because of simple physics. Whereas plutonium must be compressed with explosives to produce a nuclear explosion — a feat that is probably beyond the technical capability of terrorist groups — the process is simpler for weapons-grade HEU, which is also used in many reactors. The United States and Russia, which have supplied the world with the bulk of HEU, have stepped up efforts to secure, remove or blend these materials into low-enriched uranium (LEU), which has 20% or less of the key isotope uranium-235. Security has been upgraded at 32 facilities, and 12 countries have been declared HEU-free since 2010.

Many of these materials are located at civilian research reactors. The risks were recognized long ago; in 1978, for example, the United States began eliminating HEU fuel in these reactors. In 1992, the US Congress enacted a law requiring countries that receive its HEU to commit to converting reactors to LEU fuel. To maintain reactor performance, however, scientists needed to develop a new generation of high-density LEU fuels, which are now available for most research facilities.

This is good news, but challenges remain. Existing high-density LEU fuels cannot be used without degrading performance in 11 specialized

US and European research reactors. Certifying new fuels and converting these reactors could take nearly two decades. In January, the US National Academies of Science, Engineering, and Medicine recommended that specialized US reactors adopt an interim solution and convert to less-enriched fuel sources containing 30–45% uranium-235. This could — and should — be accomplished over several years, without impeding efforts to complete the shift to safer LEU fuel as soon as possible.

“Research reactors are just one part of the puzzle.”

Researchers also need a comprehensive strategy to maintain research reactors. The European Commission is sponsoring a research consortium called HERACLES to do just that. The White House Office of Science and Technology Policy should convene agencies and research facilities to develop a path forward, and engage internationally. Many of these specialized research reactors are getting old; in some cases, given delays with new LEU fuels, it may make sense to start anew.

But research reactors are just one part of the puzzle, and the question now is how to carry the broader nonproliferation agenda forward once Obama leaves office. His four nuclear summits have boosted political attention and accelerated progress, but the world is awash with nuclear materials. Nuclear safety and security falls to a problematic patchwork of international institutions, including the International Atomic Energy Agency (IAEA), Interpol and the United Nations, and the latest summit produced a variety of initiatives to bolster these institutions.

That is a start. Ultimately, the world needs a new convention that sets specific standards for nuclear security and allows inspections and enforcement by the IAEA. In the meantime, governments must work through existing institutions to share and implement best practices. Regardless of cost, research facilities must ensure that their nuclear materials are safe and secure. ■

Mind matters

Mental illness is moving up the global agenda — but there is still much to do.

Nominally, 2016 should be a good year for mental health. On 13 and 14 April, the World Health Organization (WHO) and the World Bank will hold an unprecedented joint conference in Washington DC to discuss mental health as both a global disease and an economic problem.

It is a welcome gesture after many snubs. Mental illness was left out of the United Nations' influential high-level meeting on non-communicable diseases (NCDs) in 2011. Almost begrudgingly, the UN gave mental health a brief mention in the entry for NCDs in its 2015 Sustainable Development Goals. As well as reducing deaths from NCDs such as heart disease by one-third, it said, the world should also “promote mental health and well-being”.

The UN failed to recognize that in terms of impact, mental disorders are at least as harmful as those better-recognized and better-funded diseases. Mental-health conditions account for 37% of the healthy life years lost to NCDs. This reflects shortened lifespans and a loss of ability to work effectively, and it translated to a global loss of US\$2.5 trillion in 2010; that figure is projected to rise to \$6 trillion by 2030 (see go.nature.com/dfdkbh). Simply ‘promoting’ mental health is not enough.

Many developed countries scarcely give mental illness more than lip service. In the United States, for instance, the topic is brought up only when a mass shooting launches a ‘conversation’ about the poor state of mental-health treatment — and then only as a distraction from the gun-control debate. Such attention rarely leads to new funding for mental-health research.

Underlying much of this neglect is the persistent bias, conscious or

not, towards believing that many mental-health conditions are a moral failing rather than the result of complex biology, despite overwhelming evidence for the latter. For poor countries wracked by infectious and childhood diseases, it can be easier to confine people with psychosis than to try to treat them. And treatment is hard to find the world over: globally, there are only nine mental-health providers for every 100,000 people, and some countries have only one or two providers. A Comment on page 25 calls for a global strategy to combat care deficiencies in both developing and developed countries.

Even in developed countries, people find it difficult to consider depression as a condition to be combated with the same clarity of purpose as heart disease. This is exacerbated by the failure to develop drugs for depression that are as clear-cut and effective as statins. This is unlikely to change soon: if anything, neuroscience is painting a bleaker picture by showing how complex these diseases are. More than 100 genetic regions have been associated with schizophrenia — and autism and depression are probably even more complex. Addressing all of these will require entirely new approaches. A News Feature on page 20 looks at the latest developments on one emerging, but still unproven, front: phone apps intended to assist people with mental-health disorders.

The annual Mental Illness Awareness Week in October promotes mental health, as do many organizations devoted to erasing the stigma and bias that harm patients and inhibit politicians. But so much more is needed in terms of improved treatments and access to care.

Finance ministers at the meeting next week should recognize the positive economic returns of investing in this direction. The UN, the WHO, the World Bank and governments should be expected to contribute something tangible to this pressing issue. The summit's attendees should come away with plans for creating specific development targets, and mechanisms for funding research, ensuring that treatment is available in low-income countries, and holding nations responsible for giving mental illnesses as prominent a place in health care as other NCDs. ■

➔ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunqv

JOHN HOULIHAN/UNIV. YORK



Recognize the value of social science

A professional body for UK social scientists can help to improve research practice — and not just in public engagement, says Andrew Webster.

If the science community is serious about integrating social science into its thinking and operations — and statements by everyone from *Nature* and the UK government to Paul Nurse, former president of the Royal Society, indicate that it is — then we social scientists must do more to make this happen.

Our input is necessary because, too often, the reach and influence of research is discovered only with hindsight. Lessons are ‘learned’ only after the social implications of new domains of science and technology have provoked controversy or challenged existing norms. Social science can help to predict these implications and plan for them. It can also help to frame science questions to make them more sensitive to economic, cultural and social factors.

Scientists have their own hands-on understanding of how the science and innovation system operates, and what demands it faces. Why should they pay attention to social-science contributions, particularly when these are not yet well known and understood across the scientific community?

Social scientists understand that many colleagues in the hard sciences are sceptical of what we can offer. We know that we need to make our contribution more widely understood and appreciated. This week, social scientists in Britain take what we hope will be a significant step.

We (including myself and colleagues Robin Williams of the University of Edinburgh and Fred Steward of the Policy Studies Institute in London) are launching a professional association, with an explicit goal of engaging with science and science policy. Researchers in the social sciences have built links with specific scientific constituencies, but have failed to engage at a more general level.

We argue that it is especially important to do this now, because all researchers are being urged to explore and make explicit how their work has reach and influence in the wider world. And we want to make clear that social science — especially science, technology and innovation studies (STIS) — should be integral to science and does not merely handle external issues, for example by addressing ‘public acceptance’.

Our new body is called the Association for Studies in Innovation Science and Technology-UK (AsSIST-UK; assist-uk.com). It has been built over 18 months and has a strongly interdisciplinary membership of more than 200 people. Unlike similar bodies in mainland Europe and the United States, which tend to look inwards to the academic discipline, it is intended to take our methods and thinking into the broader world.

One priority is to lobby for social-scientist involvement in the earliest stages of research projects, when emerging ideas are most open to discussion. We want to work at national and regional levels, from the UK government and research-funding councils to professional science bodies and the devolved

governments in the four UK nations, which are experimenting with science and technology policies.

Science and society are not discrete, as some researchers seem to assume. Knowledge — about the impacts of climate change, for example — gets its value and usefulness only when rooted in particular contexts. This makes it diverse and contested. From the perspective of a social scientist, the challenge for researchers in genomics, for instance, is not to ‘communicate’ their ideas and discoveries to a homogenous ‘public’: it is to realize that they are members of that public and act accordingly. Change in the direction of science cannot be viewed from a single position. It is relative and depends on the position of the observer, and on the interplay of diverse knowledge communities.

Social science can help in interrogating the evidence and assumptions behind theoretical models (say, for biomarkers of ageing). It can assess how technical standards are defined and applied to a new field, and how innovation shapes the way in which clinical trials are designed and conducted (as my own research has done in regenerative medicine). Models, standards and trials all rely on agreement about appropriate evidence and how it is used. This allows for fruitful discussion across the sciences and the early recognition that knowledge is provisional and may change — important if a project is to have scope for future debate.

One of the first tasks for our new group is to review the research and activities of our members, to identify their existing impacts on science (through specific projects and membership of national bodies) and how they have engaged more broadly with the wider scientific community. To increase that engagement, we aim to identify and share examples of good practice. We want to bring together STIS expertise in diverse fields such as biomedicine, energy and data analytics to inform research-council programmes — including the current move towards interdisciplinary doctoral training across the social and natural sciences.

And we want to act as a national body that can contribute to specific fields of science from their earliest days onwards. Social scientists should help to plan for the possibilities of gene editing, for example, and anticipate the challenges posed by the diversification and growth of biobanks.

To make the most of science, we must know how science operates, and understand the factors that influence it. Social scientists in the United Kingdom and elsewhere have been studying that for more than 50 years. We are ready and able to help. ■

Andrew Webster is Director of the Science and Technology Studies Unit (SATSU) at the University of York, UK.
e-mail: andrew.webster@york.ac.uk

TO MAKE
THE MOST OF
SCIENCE,
WE MUST
KNOW HOW
SCIENCE
OPERATES.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/arstjn

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

MEDICINE

Drug manufacture on demand

A fridge-sized machine can produce multiple pharmacy-quality medicines on demand in hours, and could make drug supply chains more resistant to disruption.

Klavs Jensen at the Massachusetts Institute of Technology in Cambridge and his colleagues developed a machine that can synthesize hundreds to thousands of doses of four common drugs: diazepam, diphenhydramine, fluoxetine and lidocaine. Their 'plug-and-play' system can conduct multiple-step synthesis from commercial starting materials, and can also purify and crystallize the resulting compounds into a useable product.

Such a machine could one day allow on-demand production of medicines for hospitals, humanitarian operations and pharmaceutical research. *Science* 352, 61–67 (2016)

VIROLOGY

Virus adaptation to mosquitoes

Insect-borne viruses can lose some of their vigour during their time in mosquitoes before transmission to a vertebrate host.

Gregory Ebel at Colorado State University in Fort Collins and his colleagues infected four mosquito species (including *Aedes aegypti*,



pictured) with West Nile virus, which jumps back and forth between mosquitoes and birds.

Fourteen days later, they isolated the virus from the insects' saliva and organs and sequenced its RNA. They found that viruses from different mosquito species had acquired different sets of mutations. When they put the virus into avian cells, they found that these mutations made it more difficult for the virus to survive, suggesting that it had adapted to living in the mosquito. The findings may help to understand how insect-borne viruses are

spread, the researchers say. *Cell Host Microbe* <http://doi.org/bdxr> (2016)

ASTRONOMY

White dwarf's weird atmosphere

Astronomers have discovered a white dwarf star that has an atmosphere made almost entirely of oxygen — the first of its kind to be discovered.

White dwarfs are the dense cores that get left behind when a relatively small star's nuclear fuel runs out and it sheds some of its outer layers. Usually their atmospheres

are made of hydrogen or helium, because these light elements float to the surface. By analysing light gathered by the Sloan Digital Sky Survey telescope, Kepler de Souza Oliveira Filho at the Federal University of Rio Grande do Sul in Porto Alegre, Brazil, and his colleagues identified a white dwarf that has been stripped of its light elements. This has left it with an atmosphere containing mostly oxygen, with traces of neon and magnesium.

A violent pulse of burning carbon from inside the star or a fiery merger with a companion white dwarf could



ECOLOGY

A path to better fisheries

Better-managed fisheries could yield billions of extra dollars in profit and extract millions more tonnes of fish from the oceans.

Many fish stocks are overexploited and poorly managed. Christopher Costello at the University of California, Santa Barbara, and his colleagues modelled the trajectories of 4,713 fisheries under various management systems.

Compared to a 'business-as-usual' trajectory, improved management centred on fishing rights that boost the economic value of fish could

increase fish numbers, the researchers estimate, and yield 16 million tonnes more fish by 2050. It would generate US\$53 billion in profit and there would be 619 million tonnes more biomass in the oceans as a result. The average overfished stock would recover in 10 years, and 98% of stocks would recover by 2050. Better management could improve food security, conservation and profits in nations that currently manage their fisheries poorly, say the authors. *Proc. Natl Acad. Sci. USA* <http://doi.org/bdtd> (2016)

MONTY RAKUSEN/GETTY

JAMES GATHANY/CDC

have burned off the outer layers, suggest the authors. *Science* 352, 67–69 (2016)

NEUROSCIENCE

Enzymes help to regrow nerves

A newly uncovered cell-signalling pathway could broaden strategies for repairing injured nerves.

Adult neurons in the central nervous system (CNS) cannot regenerate damaged axons, the branches that conduct electrical signals. Researchers have previously managed to induce axon regeneration in a mouse retina by deleting a gene called *Pten*, finding that the process activated the protein complex mTORC1. Bo Chen at the Yale University School of Medicine in New Haven, Connecticut, and his colleagues have now identified a second regeneration pathway. They found that deleting *Pten* in mouse retinal neurons also inactivates the metabolic enzyme GSK3 β , and that deleting this enzyme in mice with optic-nerve injuries prompted the axons to regrow.

The authors suggest that combined manipulation of the two *Pten* pathways could lead to treatments to repair injured axons in the CNS. *eLife* 5, e11903 (2016)

ANIMAL BEHAVIOUR

Hunting habits of wild dogs tracked

African wild dogs that live in woodland eschew the collaborative long-distance pursuit of prey used by their relatives on grass plains.

Alan Wilson and Tatjana Hubel at the Royal Veterinary College in Hatfield, UK, and their colleagues tracked a pack of six wild dogs (*Lycaon pictus*) in Botswana's woodland savannah, where most populations of this endangered animal now live. They found that the animals used multiple short-distance attacks in their attempts to capture prey. Unlike in dog packs on the

plains, there was no observed cooperation between the dogs other than travelling together and sharing their kills. Although only 15.5% of individual attacks resulted in a kill, sharing of food ensured that all animals ate regularly.

In another paper, the same team modelled the energy use and benefits of this hunting strategy. Whereas the low-cost, short-distance chase of wild dogs is more costly per kill than for individual hunting by cheetahs (*Acinonyx jubatus*), this is outweighed by the sharing of prey. The dogs might be better equipped to deal with different habitats than was thought, the authors say. *Nature Commun.* 7, 11033 (2016); *Nature Commun.* 7, 11034 (2016)

CANCER

Prostate cancer under attack

A nuclear-receptor protein called ROR- γ has a key role in making prostate cancer deadly.

In many cases of prostate cancer, the gene that encodes the androgen receptor becomes overactive and drives tumour progression, leading to fatal tumours that are resistant to castration therapy and contain high levels of ROR- γ . To understand its role, Hong-Wu Chen at the University of California, Davis, Yong Xu at the Guangzhou Institutes of Biomedicine and Health, China, and their team suppressed ROR- γ in cancerous cell lines and in human tumours transplanted into mice. This suppression blocked expression of the androgen-receptor gene and inhibited tumour growth, suggesting that ROR- γ is a crucial component of the castration-resistant tumours.

Current therapies target androgen-receptor levels, but mutations in tumours often make them resistant to such treatment. Targeting ROR- γ could help patients to fight advanced prostate cancer. *Nature Med.* <http://doi.org/bdtk> (2016)

ARCHAEOLOGY

Neanderthal uncovered

With the help of a protein-fingerprinting technique, researchers have identified one small chunk of bone out of thousands as belonging to a Neanderthal.

Bone fragments from archaeological sites are often too small to identify the species of origin. Samantha Brown at the University of Oxford, UK, and her colleagues analysed 2,315 such pieces from the Denisova Cave in southern Siberia — where Denisovans (an archaic human group) were first identified.

The team extracted proteins from each fragment, finding human collagen protein in only one. Mitochondrial DNA sequencing and radiocarbon dating then showed that this 1.7-gram piece was from a Neanderthal — not a Denisovan — living more than 50,000 years ago.

Sci. Rep. 6, 23559 (2016)

PALAEOONTOLOGY

Fossil snake colours revealed

Fresh evidence of coloration in ancient animals has been discovered in the 10-million-year-old fossilized remains of a snake.

The colours of many extinct animals, including early birds, reptiles, mammals and amphibians, have been reconstructed through evidence of melanin pigments preserved in fossils. Maria McNamara at University College Cork, Ireland, and her colleagues now report evidence of coloration created by a different type of pigment called a carotenoid, preserved in minerals that replaced the tissues during fossilization.

The team discovered pigment cells, called chromatophores, that would have contained carotenoids and light-reflecting granules in the skin of the fossilized snake,

which was discovered near Teruel in Spain. The findings suggest that the animal had a pale belly and a green back and sides, with brown-black and yellow-green blotches — possibly a form of camouflage. The coloration could also have been a way to signal to other snakes.

Curr. Biol. <http://doi.org/bdxx> (2016)

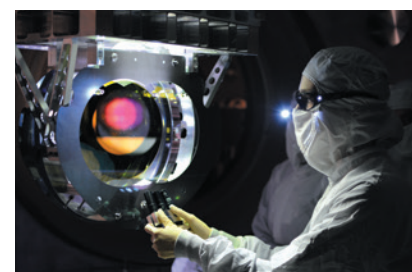
ASTROPHYSICS

Black-hole crackle-and-pop

Gravitational 'noise' could be easier to detect than previously thought.

Gravitational waves were first observed in 2015 by the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors (**pictured**). This demonstrated the existence of binary black holes and suggested that these are more abundant than some previous estimates indicated. Tania Regimbau, at the Cote d'Azur Observatory in Nice, France, and her colleagues from the LIGO/Virgo collaboration calculated the odds of detecting a constant background noise made by similar black-hole mergers across the observable Universe that are too distant to be detected individually. They find that LIGO and Virgo could potentially detect this background when they reach peak sensitivity in about four years time.

Phys. Rev. Lett. 116, 131102 (2016)



➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

PEOPLE

Fraud punished

A Parkinson's disease researcher in Australia pleaded guilty to research fraud and was handed a two-year suspended prison sentence by a court in Brisbane on 31 March. Bruce Murdoch, formerly of the University of Queensland in Brisbane, was found to have falsified results published in the *European Journal of Neurology* in 2011; three of his papers have been retracted. In a statement to the blog Retraction Watch, University of Queensland vice-chancellor Peter Høj said that the university had reimbursed around Aus\$175,000 (US\$132,000) to funding bodies associated with Murdoch's work.

EVENTS

Emergency over

The Ebola outbreak that has killed more than 11,000 people in Guinea, Liberia and Sierra Leone is no longer a public-health emergency, said the World Health Organization (WHO) in Geneva, Switzerland, on 29 March. The WHO noted that the virus is now unlikely to spread internationally, and that the affected countries have the capacity to deal with new cases. A campaign in Guinea last month administered an experimental Ebola vaccine to nearly 800 people who may have come into contact with 8 individuals with the virus. Liberia has recorded two new cases since the announcement: a woman who died on 31 March and her five-year-old son.

ET search starts

The SETI Institute in Mountain View, California, has kicked off a search for signals from extraterrestrial civilizations that might be living on



EPA EUROPEAN PRESS PHOTO AGENCY B.V./ALAMY STOCK PHOTO

Ice wall to stem Fukushima leak

The Tokyo Electric Power Company (TEPCO) on 31 March began freezing the soil surrounding reactors 1 to 4 of the disaster-stricken Fukushima Daiichi nuclear power plant. A refrigeration system (pictured) is creating a 30-metre deep, 1.5-kilometre-long wall of frozen ground that aims to stop groundwater from flowing under

the plant and carrying radioactive isotopes into the sea. More than 300 tonnes of water per day are pumped into the damaged reactors to stop their cores from overheating and contributing to the radioactivity leak. TEPCO expects that it will take months for the ¥35-billion (US\$316-million) project to seal the zone.

planets orbiting any of the 20,000 nearest red dwarf stars. Red dwarfs are dimmer and cooler than the Sun, but they make up the bulk of stars in the Galaxy, increasing the odds of finding life there. The two-year search will be conducted at the Allen Telescope Array in northern California, the institute announced on 30 March.

China's Go

A team of Chinese scientists plans to challenge Google DeepMind's Go-playing artificial-intelligence algorithm with its own program by the end of 2016, Chinese state news has reported. The DeepMind program, known as AlphaGo, beat a leading human player, South Korea's Lee Sedol, by four games to one in March. Reporting from an event

organized by the Chinese Go Association and the Chinese Association for Artificial Intelligence, *Shanghai Securities News* said on 31 March that a team from China will issue the challenge by the end of the year.

Science in space

Two scientific payloads travelled to the edge of space on 2 April in the latest test of Blue Origin's commercial space vehicle, New Shepard. Amazon founder Jeff Bezos's reusable spacecraft took off from a Texas launch site, flew to an altitude of 103 kilometres and successfully landed 11 minutes later. On board was an experiment from the University of Central Florida in Orlando investigating how dust and rubble settle in microgravity, and the 'box of rocks experiment' from the

Southwest Research Institute in San Antonio, Texas, to work out how regolith forms on asteroids.

RESEARCH

Crater drilling

A two-month expedition to drill into the 200-kilometre-wide Chicxulub crater, which formed 66 million years ago when an enormous asteroid smashed into the planet, began on 1 April. The aftermath of the impact obliterated most life on Earth, including the dinosaurs. From a drill-ship off the coast of Yucatán, Mexico, researchers will start to penetrate one of Chicxulub's most striking features — its 'peak ring', a circle of mountains within the crater floor. Scientists have yet to fully explain how peak rings form, even though they

are common in big impact craters across the Solar System. See go.nature.com/pgxb18 for more.

Gorilla decline

Numbers of the largest primate on the planet, Grauer's gorilla (*Gorilla beringei graueri*), have plummeted since 1995, according to a report from the Wildlife Conservation Society (WCS). The report, published on 4 April, says that the numbers have dropped from an estimated 17,000 in 1995 to 3,800 today, a 77% decrease. Grauer's gorillas live in the wild only in the eastern Democratic Republic of the Congo, and the WCS report blames the decline on illegal hunting around mining sites, the civil war in the country from 1996 to 2003, and habitat loss.

POLICY

Nuclear security

More than 50 countries, most represented by their heads of state, made a variety of commitments intended to prevent nuclear terrorism at the conclusion of a nuclear summit in Washington DC on 1 April. The meeting was the fourth biennial summit in a process initiated in 2009 by US President Barack Obama (pictured with Canadian Prime Minister Justin Trudeau). Much of the focus has been on reducing civilian



stocks of highly enriched uranium at research reactors. At least 28 reactors have been shut down or converted to low-enriched uranium since 2009, but challenges remain in converting 11 high-performance research reactors. See page 5 for more.

BUSINESS

Patent pledge

Drugmaker GlaxoSmithKline (GSK) has announced plans to improve access to its medicines in the world's poorest countries. The company said on 31 March that it would stop filing drug patents in many developing countries. That means that generic manufacturers in those nations would be able to supply copycat versions of GSK's drugs without worrying about lawsuits. GSK also signalled that it intended to improve access to low-cost drugs that can help to address the growing burden of cancer in the developing world.

Public-health advocates have embraced the news and are urging other drug companies to follow suit. See go.nature.com/nqhggi for more.

FACILITIES

Laser beam added

The Linac Coherent Light Source (LCLS), the world's brightest X-ray free-electron laser, began a US\$1-billion construction project on 4 April to add a second beam. LCLS-II, based at the SLAC National Accelerator Laboratory in Menlo Park, California, will accelerate electrons through superconducting niobium cavities to produce X-ray pulses 10,000 times more concentrated and firing 8,000 times faster than X-rays produced by the \$414-million LCLS, which started operations in 2009. This will enable it to image processes that occur on smaller scales and faster timescales. Construction will last until the early 2020s.

FUNDING

Career boost

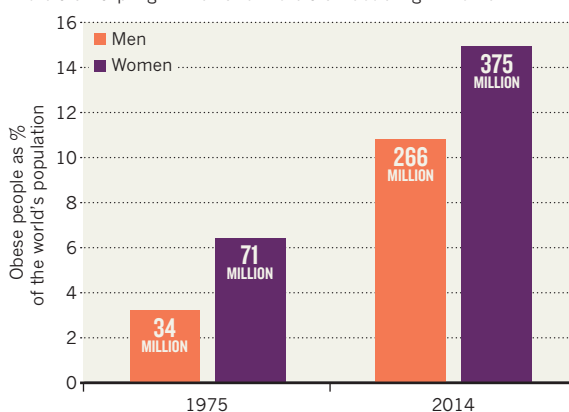
Four philanthropic organizations have created an international research programme focused on early-career scientists. Announced on 29 March, the International Research Scholars Program will select up to 50 members who are not originally from one of the G7 countries, but have trained in the United States or the United Kingdom for at least one year. Awardees will each receive a total of US\$650,000 over five years. The sponsors are: the Howard Hughes Medical Institute in Chevy Chase, Maryland; the Bill & Melinda Gates Foundation in Seattle, Washington; the London-based Wellcome Trust; and Lisbon's Calouste Gulbenkian Foundation.

TREND WATCH

More of the global population is now obese than is underweight, according to a study of 186 countries from 1975 to 2014 (see *Lancet* 387, 1377–1396; 2016). The proportion of obese men more than tripled and that of obese women more than doubled during that period. Many people are still underweight in the world's poorest regions, particularly in Asia and Africa. But the global average weight of a person grew by 1.5 kilograms each decade. See go.nature.com/ysliff for more.

OBESITY RISE OVER 40 YEARS

Data from 1975 to 2014 show that obesity has risen dramatically, more than tripling in men and more than doubling in women.



CORRECTION

The item 'Boaty McBoatface' (*Nature* 531, 419; 2016) wrongly described NERC as the 'National' instead of the Natural Environment Research Council.

NATURE.COM

For daily news updates see: www.nature.com/news

NEWS IN FOCUS

COSMOLOGY Physicists prepare to test enigmatic dark-matter claim **p.14**

ZIKA Outbreak highlights role for controversial fetal tissue research **p.16**

BIODIVERSITY Negotiations begin to protect high seas **p.18**



MENTAL HEALTH Apps for the brain proliferate despite little proof they work **p.20**

ERIC THAYER/NYT/REDUX/EVINE



US Republican presidential candidate Donald Trump has proposed building a wall between the United States and Mexico.

POLITICS

Trump's immigration stance stokes fears for science

Rhetoric in US presidential campaign concerns researchers — particularly Muslims.

BY HEIDI LEDFORD

Razi Nalim has lived in the United States for 30 years. An engineer at Indiana University–Purdue University Indianapolis, he often travels around the world to recruit science and engineering students to his university. But last week, on the cusp of a recruitment trip to India, he hesitated when asked whether he would still encourage foreign, Muslim students to work or study in the United States.

“I would still say the opportunity for doing

cutting-edge science here is unmatched,” said Nalim, who is Muslim. “Where I think I would caution people to think more carefully is longer term: where would they want to live and raise a family? That’s a harder question to answer.”

For Nalim and others, the roots of such concerns are apparent. In December, US presidential candidate Donald Trump, who has campaigned against immigration, boasted that he would ban Muslims from entering the country if elected. (On 30 March, Trump — now the Republican front runner — said that

he would make exceptions for some Muslims, notably his wealthy Muslim friends.)

Science advocates worry that Trump’s broader anti-immigration stance could pose a threat to US research dominance. Roughly 5% of all students in the United States hail from other countries — including more than 380,000 people studying science, engineering, technology or mathematics. “We’ve always been a nation which has welcomed scientific brainpower from other countries,” says Mary Woolley, president of Research!America, a ▶

► science-advocacy group in Alexandria, Virginia. “We don’t want that to turn around now.”

Scientific issues have scarcely been mentioned on the campaign trail so far. Hillary Clinton, the Democratic front runner, has pledged to boost support for research into Alzheimer’s disease, and has pushed back against Trump’s anti-immigration and anti-Muslim stance. When she was a senator, Clinton backed health and research-related bills, and as first lady to former president Bill Clinton, she advocated for research on women’s health.

Trump is a wealthy real-estate mogul with no political legacy to mine for clues as to his scientific opinions. In the course of the campaign, he has linked autism to childhood vaccines, and dismissed climate change. (“It’s called weather,” he said.) In October, conservative radio host Michael Savage suggested on air that if elected, Trump should appoint him as head of the US National Institutes of Health (NIH). “Well, you know you’d get common sense if that were the case, that I can tell you,” Trump replied, during the light-hearted conversation. “Because I hear so much about the NIH, and it’s terrible.”

With little more than this to go on, advocates of science funding are worried. “It feels like there’s a lot of cynicism toward science and scientists, and that’s concerning,” says Benjamin Corb, public-affairs director at

the American Society for Biochemistry and Molecular Biology in Rockville, Maryland.

Trump’s position on immigration is clearer. He frequently boasts that if elected, he would build a wall along the border with Mexico — and force Mexico to pay for it — which has earned him both supporters and derision. A

“There’s a lot of cynicism toward science and scientists.”

President Trump could bode ill for long-running efforts to boost the number of foreign professionals working in the United States on visas for highly skilled workers, known as H-1Bs. But Trump’s statements regarding H-1B visas have been difficult to parse. At times, he has advocated bringing skilled workers into the country; at others, he has said that the H-1B programme is too often abused and should be restricted.

Such statements worry Brad Hayes, a computer scientist at the Massachusetts Institute of Technology in Cambridge. Hayes is an US citizen, but says that some of his most outstanding colleagues are not. “A lot of them want to end up here after they get their PhDs, but now that’s in doubt,” he says. “We absolutely want these people to stay. If they get lumped in with this ‘close our borders, keep everybody out,’ we’re doing ourselves a disservice.”

Hayes inadvertently cast a spotlight on the simplicity of Trump’s rhetoric when he decided to use a neural network to model Trump’s noticeably repetitive and simplistic speech patterns. He has been posting the results — computer-generated parody quotes based on Trump’s campaign speeches — on Twitter using the handle @DeepDrumpf. (Trump’s ancestral name, Drumpf, was changed by the family several generations ago.)

“We’re going to build the wall,” says one tweet, in reference to Trump’s Mexico plan. Hayes says that the project was only meant to be fun, but it ended up making a point. “A lot of the rhetoric that’s being used is fairly content-light.”

But that rhetoric is having an effect, says Ehab Abouheif, a developmental biologist at McGill University in Montreal, Canada, who is Muslim. On a recent trip to be interviewed for a position in the United States, recruiters’ “constant question was, ‘Are you really sure you would want to come?’” he says. “My scientist colleagues are really scared.”

To Abouheif, who fondly remembers completing his PhD and his postdoc in the United States, the current climate is surreal. “If you are trying to stop Muslims from coming in, it means that the ones who are there already are not going to feel comfortable either,” he says. “It would be a shame to alienate this big swathe of society.” ■

COSMOLOGY

Controversial dark-matter claim faces ultimate test

Multiple teams finally have the material they need to repeat enigmatic experiment.

BY DAVIDE CASTELVECCHI

It is the elephant in the room for dark-matter research: a claimed detection that is hard to believe, impossible to confirm and surprisingly difficult to explain away. Now, four instruments that will use the same type of detector as the collaboration behind the claim are in the works or poised to go online. Within three years, the experiments will be able to either confirm the existence of dark matter — or rule the claim out once and for all, say the physicists who work on them.

“This will get resolved,” says Frank Calaprice of Princeton University in New Jersey, who leads one of the efforts.

The original claim comes from the DAMA collaboration, whose detector sits in a laboratory deep under the Gran Sasso Massif, east of Rome. For more than a decade, it has reported

overwhelming evidence¹ for dark matter, an invisible substance thought to bind galaxies together through its gravitational attraction. The first of the new detectors to go online, in South Korea, is due to start taking data in a few weeks. The others will follow over the next few years in Spain, Australia and, again, Gran Sasso. All will use sodium iodide crystals to detect dark matter, which no full-scale experiment apart from DAMA’s has done previously.

Scientists have substantial evidence that dark matter exists and is at least five times as abundant as ordinary matter. But its nature remains a mystery. The leading hypothesis is that at least some of its mass is composed of weakly interacting massive particles (WIMPs), which on Earth should occasionally bump into an atomic nucleus.

DAMA’s sodium iodide crystals should produce a flash of light if this happens in the

detector. And although natural radioactivity also produces such flashes, DAMA’s claim to have detected WIMPs, first made in 1998, rests on the fact that the number of flashes produced per day has varied with the seasons.

This, they say, is exactly what is expected if the signal is produced by WIMPs that rain down on Earth as the Solar System moves through the Milky Way’s dark-matter halo². In this scenario, the number of particles crossing Earth should peak when the planet’s orbital motion lines up with that of the Sun, in early June, and should hit a low when its motion works against the Sun’s, in early December.

There is one big problem. “If it’s really dark matter, many other experiments should have seen it already,” says Thomas Schwetz-Mangold, a theoretical physicist at the Karlsruhe Institute of Technology in Germany — and none has. But at the same time, all attempts to

find weaknesses in the DAMA experiment, such as environmental effects that the researchers had not taken into account, have failed. “The modulation signal is there,” says Kaixuan Ni at the University of California, San Diego, who works on a dark-matter experiment called XENON1T. “But how to interpret that signal — whether it’s from dark matter or something else — is not clear.”

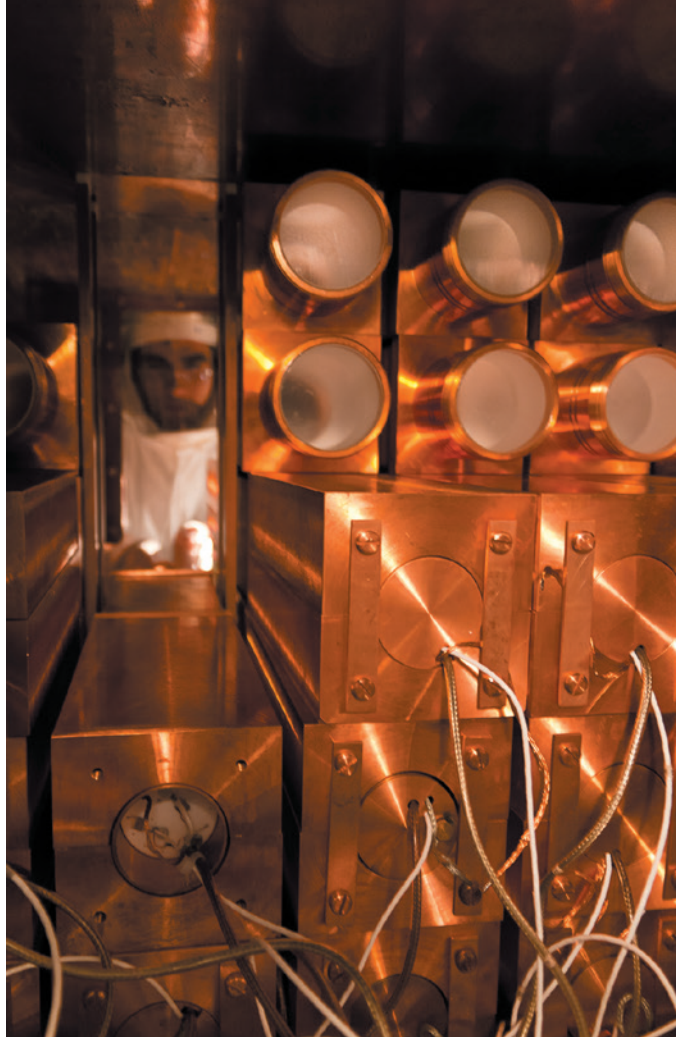
No other full-scale experiment has used sodium iodide in its detector, although the Korea Invisible Mass Search (KIMS), in South Korea, used caesium iodide. So the possibility remains that dark matter interacts with sodium in a different way to other elements. “Not until someone has turned on a detector made of the same material will you grow convinced that nothing is there,” says Juan Collar at the University of Chicago, Illinois, who has worked on several dark-matter experiments.

Many have found it challenging to grow sodium iodide crystals with the required purity. Contamination by potassium, which has a naturally occurring radioactive isotope, is a particular problem.

But now three dark-matter-hunting teams — KIMS; DM-Ice, run from Yale University in New Haven, Connecticut; and ANAIS, at the University of Zaragoza, Spain — have each obtained crystals with about twice the level of background radioactivity of DAMA’s. That is pure enough to test its results, they say.

The KIMS and DM-Ice teams have built a sodium iodide detector together at Yangyang Underground Laboratory, 160 kilometres east of Seoul. This instrument uses an ‘active veto’ sensor that will enable it to separate the dark-matter signal from the noise better than DAMA does, says Yeongduk Kim, the director of South Korea’s Center for Underground Physics in Daejeon, which manages KIMS.

ANAIS is building a similar detector in the Canfranc Underground Laboratory in the Spanish Pyrenees. Together, KIMS/DM-Ice and ANAIS will have about 200 kilograms of sodium iodide, and they will pool their data.



The DAMA team uses sodium iodide housed in copper to hunt for dark matter.

That is comparable to DAMA’s 250 kilograms, enabling them to catch a similar number of WIMPs, they say. Even though the newer detectors will have higher levels of background noise, it should still be possible to either falsify or reproduce the very large DAMA signal, says Reina Maruyama of Yale, who leads DM-Ice.

But Calaprice argues that high purity is more important than mass. He and his collaborators have developed a technique to lower contamination, and in January announced that they were the first to obtain crystals purer than DAMA’s. He expects to reduce the background levels further, to one-tenth of DAMA’s.

The project, SABRE (Sodium-iodide with Active Background Rejection), will put one detector at Gran Sasso and the other at the Stawell Underground Physics Laboratory, which is being built in a gold mine in Victoria, Australia. SABRE will also use a sensor to pull

out the dark-matter signal from noise, and will have a total mass of 50 kilograms.

SABRE should complete its research and development stage in about a year, and will build its detectors soon after that, says Calaprice. It will then make its technology available to other labs — something that DAMA did not do. And having twin detectors in both the Northern and Southern hemispheres could clarify whether environmental effects could have mimicked dark matter’s seasonality in DAMA’s results — if the signal is from WIMPs, then both detectors should see peaks at the same time.

DAMA will wait at least until 2017 to release its latest results, says spokesperson Rita Bernabei of the University of Rome Tor Vergata. She is not holding her breath about the upcoming sodium iodide detectors. “Our results have already been verified in countless cross-checks in 14 annual cycles, so we have no reason to get excited about what others may do,” she says. If other experiments do not see the annual modulation, she adds, her collaboration will conclude that they do not yet have

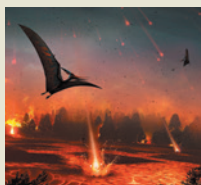
sufficient sensitivity.

Could the teams prove DAMA right? “I was unwilling to believe the DAMA results or even take them seriously at first,” says Katherine Freese, a theoretical astroparticle physicist at the University of Michigan in Ann Arbor, who with her collaborators first proposed the seasonal modulation technique used by DAMA². But, in part because of a lack of any other explanation for their signal, she is now more hopeful. The fact that many have tried and failed to repeat DAMA’s experiment shows that it is not easy, says Elisabetta Barberio at the University of Melbourne, who leads the Australian arm of SABRE. “The more one looks into their experiment, the more one realizes that it is very well done.” ■

1. Bernabei, R. *et al.* *Eur. Phys. J. C* **73**, 2648 (2013).
2. Drukier, A. K., Freese, K. & Spergel, D. N. *Phys. Rev. D* **33**, 3495–3508 (1986).



TOP STORY



Geologists to drill into heart of dinosaur-killing impact go.nature.com/pgxb18

MORE NEWS

- Biology software promises easier way to program living cells go.nature.com/yzxg3b
- How human sacrifice propped up the social order go.nature.com/msznr
- Did humans drive ‘hobbit’ species to extinction? go.nature.com/whuehv

NATURE PODCAST



Mental-health apps; ritual human sacrifice; and supernova debris on Earth. nature.com/nature/podcast



Researchers are trying to establish whether Zika virus causes the birth defect microcephaly.

PUBLIC HEALTH

Zika highlights role of fetal-tissue research

Controversial tissue studies could prove crucial to probing link between virus and birth defects.

BY ERIKA CHECK HAYDEN

A protein that helps Zika virus infect adult skin cells might also give the virus access to stem cells that make brain cells, suggests a study carried out on donated human fetal tissue.

The result — published on 30 March in *Cell Stem Cell*¹ — is part of a growing body of research that seeks to determine how Zika might cause birth defects, but that requires a type of tissue that is increasingly controversial for researchers in the United States.

Recent advances in neuroscience and cell technology have given hints as to why some babies born to Zika-infected mothers have abnormally small heads — a condition called microcephaly — and other birth defects. But to fully understand what is happening in the womb, some scientists say that they need to study tissue from fetuses, which can be donated by couples who terminate pregnancies.

Researchers already knew that a protein called AXL enabled Zika to enter human skin cells. Now, Arnold Kriegstein, a neuroscientist at the University of California, San Francisco,

(UCSF) and his colleagues show that the protein is also made by cells in the fetus that form the eyes and the brain. AXL could provide a means for Zika virus to infect these cells.

Two other studies published this month^{2,3} showed that Zika specifically targets and

“Many fewer people are willing to donate, and it’s slowing us down.”

kills neuron-forming cells, including those in organoids — brain-like structures derived from reprogrammed human skin cells. These studies suggest that Zika causes microcephaly by damaging fetal cells that make the brain, says neuroscientist Patricia Pestana Garcez of the Federal University of Rio de Janeiro, Brazil, who led one of the studies.

Kriegstein’s study used fetal tissue donated by patients treated at UCSF medical facilities. But such material may get harder to come by, because the collection and use of fetal cells is under renewed scrutiny in the United States. Last July, an anti-abortion group called the Center for Medical Progress in Irvine, California, released video of employees from

the non-profit health-care provider Planned Parenthood discussing the sale of fetal tissue from abortions for research. Members of the US House of Representatives are now investigating the use of fetal tissue in research.

Scientists in the United States worry that the controversy could hamper essential research on the Zika virus. “Many fewer people are willing to donate, and it’s slowing us down,” says Susan Fisher, a stem-cell and developmental biologist at UCSF.

Fisher is studying how Zika virus is transmitted from mother to baby. She has found AXL in fetal cells called trophoblasts that anchor the placenta, which supplies a fetus with blood and nutrients, to the mother’s uterus. These cells are known to transmit infections such as cytomegalovirus from mother to baby. “This suggests that the placenta is extremely capable of transmitting Zika,” says Fisher, whose studies rely on fetal tissue donated from terminated and full-term pregnancies.

Carolyn Coyne, a virologist at the University of Pittsburgh in Pennsylvania, says that fetal tissue is particularly crucial for studies of Zika because the virus seems to be able to harm a fetus throughout pregnancy⁴. “It is absolutely essential to study Zika infection in human fetal tissue,” says Coyne. “These types of studies need to extend to all stages of pregnancy.”

Because abortion is illegal or highly restricted in many Latin American countries, laboratory research on neural development in the regions hit hardest by Zika relies mainly on other types of human tissue, such as organoids. Researchers in Brazil, for example, are studying the lethality of different Zika viruses in neurons and organoids derived from cord blood.

Both Fisher and Kriegstein are planning further studies to test how Zika infects developing brain and placental cells. They argue that such studies are crucial to establish why the virus damages babies’ brains, and whether this can be prevented.

The scientists will also use organoids and animal models, but they note that neither of these is a perfect substitute for human fetal tissue. For instance, researchers aren’t sure how faithfully the growth of brain organoids replicates human brain development. “It’ll be important to demonstrate in human tissue exactly how the virus is creating the pattern of damage that is emerging,” Kriegstein says. “In situations like this, where there’s considerable time pressure to try to unravel what’s going on and to protect the developing human brain, it’s especially important.” ■

1. Nowakowski, T. J. *et al. Cell Stem Cell* <http://dx.doi.org/10.1016/j.stem.2016.03.012> (2016).
2. Garcez, P. P. *et al. PeerJ Preprints* **4**, e1817v3 (2016).
3. Tang, H. *et al. Cell Stem Cell* <http://dx.doi.org/10.1016/j.stem.2016.02.016> (2016).
4. Brasil, P. *et al. N. Eng. J. Med.* <http://dx.doi.org/10.1056/NEJMoa1602412> (2016).

HYDROLOGY

Snow sensors seek best way to track the white stuff

Airborne experiments aim to fill in the blanks of global water resources as the climate changes.

BY ALEXANDRA WITZE

When it comes to monitoring the world's frozen places, ice gets most of the love. Satellites such as CryoSat-2, run by the European Space Agency (ESA), measure minute changes in Earth's melting ice sheets. Now another group of cryospheric scientists hopes to get in on the action — by monitoring not ice, but snow.

Snow measurements are crucial for understanding the world's water resources. But observations lag behind those of ice, mainly because remote sensing doesn't work consistently across all snowy environments. From mountains to prairies to tundra, the sheer variety of landscapes has frustrated efforts to produce high-resolution, worldwide maps.

"The biggest hole in our knowledge of the global water budget is snow," says Jeffrey Deems, a research scientist at the US National Snow and Ice Data Center in Boulder, Colorado. "We really have no idea how much is out there."

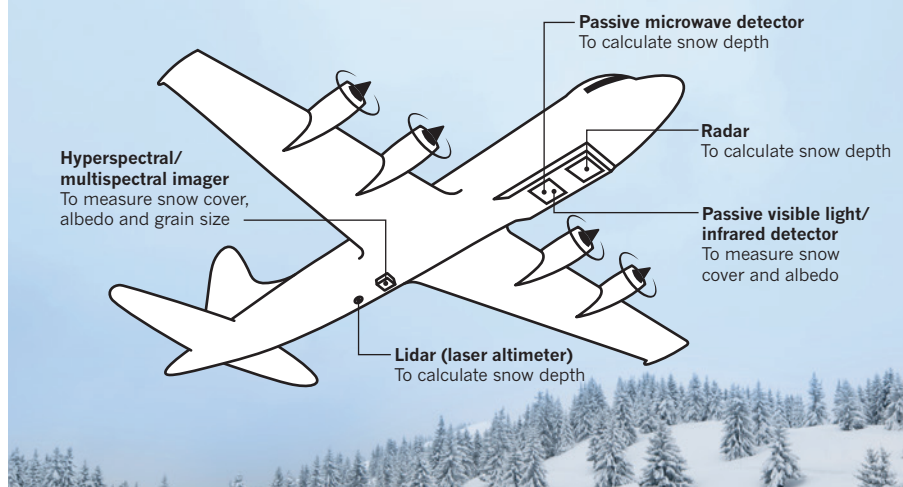
Last week, at a workshop in Seattle, Washington, Deems and his colleagues settled on a plan to change that, when they laid out details for a multiyear NASA field campaign scheduled to begin in September. The SnowEx experiment will fly aeroplanes carrying a combination of remote-sensing instruments — including radar, laser altimeters (lidar) and multispectral imagers — over snowy landscapes. The goal is to see which techniques work best for studying snow, and to combine those in a design for a snow-sensing satellite.

Snow information is becoming more crucial as the climate changes, says Matthew Sturm, a snow scientist at the University of Alaska Fairbanks. More than 1.2 billion people worldwide rely on mountain snowpacks for water — but in many areas, snowfall may decrease in the future (J. S. Mankin *et al. Environ. Res. Lett.* 10, 114016; 2015). In California, for example, the ongoing drought means that water managers are increasingly eager for any information about how much runoff to expect, and when, throughout the summer.

In Alaska, changing snow cover affects how fast permafrost thaws, destabilizing the landscape. And as Arctic sea-ice cover shrinks, so too does its protective snow cover, leading to feedback loops of increasing ice destruction.

EYES ON THE SNOW

Remote-sensing measurements could finally let scientists monitor Earth's snow resources — which provide drinking water for billions of people. NASA is planning to test various combinations of sensors to see which do best at quantifying how much snow lies on a landscape and how quickly it is likely to melt away.



SOURCE: NASA; PHOTO: MMAC72/GETTY

Current satellites have limited ability to track these changes. ESA's now-concluded GlobSnow project used satellite microwave data to map global 'snow water equivalent' — the crucial estimate of how much water is contained in the world's snowpacks, calculated by multiplying snow depth by density. But GlobSnow's maps, with pixels 25 kilometres on each side, are too low-resolution for precise estimates in individual watersheds.

In the past few years, NASA and ESA have each rejected proposals for more-detailed snow-observing satellites. Both missions would have used radar to measure snow depth and calculate snow water equivalent, and both were doomed by doubts that researchers could reliably extract that information from the type of radar proposed. "For a long time we were on a quest for a single sensor," Sturm says. "The snow's just not that simple."

Now the focus is shifting to testing multiple sensors simultaneously, to see which combination works best. The first SnowEx flights will carry lidar instruments and several types of radar (see 'Eyes on the snow') over the Sierra Nevada or Rocky Mountains in western North America. Each will measure snow cover by monitoring how lidar or radar pulses bounce off the ground and reflect back to the plane. The instruments will include new technologies that,

when used together, may avoid the problems of the rejected satellite radar, says Edward Kim, lead scientist for SnowEx at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

One method is already becoming enormously popular in California and other parts of the western United States. For the past several years, NASA has flown a lidar-equipped plane called the Airborne Snow Observatory (ASO) over several western watersheds. The observatory measures the shape of the terrain in the summer, when there is no snow on the ground, and then returns throughout the winter to measure the changing depth of the snowpack.

Project scientists can build up highly detailed maps — down to 1.5-metre resolution — of watersheds such as the Tuolumne River Basin, which supplies the city of San Francisco in California. Water managers use the resulting data to estimate how much runoff to expect in the spring. "We've never had that across these mountain basins before," says Thomas Painter, who heads the ASO project at NASA's Jet Propulsion Laboratory in Pasadena, California.

SnowEx will build on the ASO by testing extra instruments. "This is the number one priority," says Jessica Lundquist, a hydrologist at the University of Washington in Seattle. "We need to figure out how to measure snow right." ■

NEUROSCIENCE

Human Brain Project releases computing tools

Move signals effort to benefit wider community.

BY QUIRIN SCHIERMEIER & ALISON ABBOTT

Europe's major brain-research project has unveiled a set of prototype computing tools and called on the global neuroscience community to start using them.

Release of the tools, which marks the start of the operational phase of the Human Brain Project (HBP), could help to allay concerns about the €1-billion (US\$1.1-billion) venture's benefits to the wider scientific community. But it is not yet clear how the platforms will resonate with researchers outside the project.

"At this point, no one can say whether or not the research platforms will be a success," says Andreas Herz, chair of computational neuroscience at the Ludwig Maximilian University of Munich in Germany.

The computing platforms include brain-simulation tools, visualization software and a pair of remotely accessible supercomputers that researchers can use to study brain processes in real time. Some will be freely accessible, others available only on the success of a peer-reviewed application (see go.nature.com/r1pjpz).

The HBP's general mission is to combine different types of neuroscience data to reconstruct the human brain computationally at different scales, and to simulate it. But the project has attracted controversy. In 2014, around 150 neuroscientists from outside the project signed a petition claiming that it was running off its scientific course, and pledged to boycott it unless their concerns were addressed. An independent review completed in March 2015 stressed, among other things, that the computing infrastructure created by the HBP must be useful to the wider scientific community.

Release of the neuroscience tools is a sign that the HBP's leadership accepts that the project needs to focus on providing concrete services for the neuroscience community at large, says Herz. But he cautions that the project still hinges on "logical flaws", such as "the dream" that sparse recordings from neurons can generate dense data.

The platforms are to be developed with a view to eventually becoming part of a permanent, pan-European research infrastructure, say the HBP's organizers — a step that would require the project to secure permanent funding commitments from governments. ■



DAVID FLEETHAM/VISUALS UNLIMITED/CORBIS

The high seas host a wealth of marine life but lack comprehensive rules to guide conservation.

MARINE BIOLOGY

Talks aim to tame marine Wild West

Nations debate how to protect biodiversity in the high seas.

BY DANIEL CRESSEY

They might host habitats of huge ecological importance, but two-thirds of the world's oceans lie beyond the authority of national governments. On 28 March, members of the United Nations began negotiating the first global treaty to impose conservation and sustainability on the high seas. As well as vastly increasing the areas that could be set aside to protect endangered species, a legally binding treaty could usher in laws on the use of marine organisms in the search for drugs and cosmetics.

"This is something that could change the future of how we manage our oceans," says Elizabeth Wilson, director of international ocean policy at the Pew Charitable Trusts in Washington DC, which has advocated for such a treaty.

Currently, nations can claim waters up to 200 nautical miles (370 kilometres) from their shorelines as 'exclusive economic zones'; everything beyond these is designated the

high seas (see 'Sea change'). There are treaties that govern specific high-seas activities, such as mining the sea bed for minerals or laying cables, and agreements that regulate some kinds of fishing in various areas. But there is no comprehensive set of regulations on biodiversity and conservation.

The ongoing treaty discussions, which are at the UN headquarters in New York, are the first talks of a preparatory committee, known as Prep Com — and the result of years of wrangling between member nations. If they, and three more meetings before the end of 2017, are successful, a text could be voted on by UN members as soon as 2018, Wilson says.

One goal of the UN negotiations is a mechanism for the creation of marine protected areas (MPAs) on the high seas. As part of the UN Convention on Biological Diversity, most governments are now committed to designating at least 10% of the oceans as protected areas by 2020 — currently, around 2% lies in MPAs. But legally, they can create MPAs only in their own waters, so some of

the most ecologically important areas remain out of reach.

Moreover, in a study published last month, marine scientists Callum Roberts, Bethan O’Leary and their colleagues at the University of York, UK, conclude that the 10% target is much too low. On the basis of a review of 144 studies that assessed the adequacy of the UN target, they estimate that more than 30% of the ocean must be protected to achieve goals such as protecting biodiversity and minimizing the risk that fish populations will collapse (B. C. O’Leary *et al. Conserv. Lett.* <http://doi.org/bdxw>; 2016). But it would be near impossible to set aside this much-larger portion without being able to create MPAs in the high seas, notes Roberts. “These negotiations over the next two years are going to be vitally important,” he says.

BROAD APPROACH

Others say that MPAs aren’t necessarily the best way to protect the oceans, because they can lead to problems such as overfishing in non-protected areas. Rather than setting aside large areas for protection, a better approach would be to manage activities properly across the ocean, says oceanographer James Cowan

SEA CHANGE

Negotiations have begun to allow the creation of marine reserves not just in the exclusive economic zones that can be claimed by nations, but throughout the oceans.



at Louisiana State University in Baton Rouge.

Another issue being discussed during the negotiations is how to govern the search for genetic resources — plants and animals that could yield products such as drugs or cosmetics.

Bioprospecting is on the rise in the oceans, particularly in the high seas, says Glen Wright, a marine-policy researcher at the Institute for Sustainable Development and International Relations in Paris.

The G77 group of developing nations argues that genetic resources from the high seas are part of the ‘common heritage of mankind’, so any profits from them should be shared among all nations. Other groups, including the European Union would rather avoid such a formal status, says Wright,

and focus instead on a practical mechanism for sharing the benefits. But at this stage, say Wright and Wilson, it is unclear what rules might be brought in, and what mechanisms might be used to enforce them. ■

CORRECTIONS

In the News story ‘Clash over killer-whale captivity’ (*Nature* **531**, 426–427; 2016), we erroneously referred to Jeffrey Ventre as a veterinary surgeon. He is, in fact, a physician. And in the story ‘Scientists say “no” to UK exit from Europe in *Nature* poll’ (*Nature* **531**, 559; 2016), the labels numbered 2 and 4 in the graphic ‘What impact would a UK exit from the EU have on UK science?’ were the wrong way round. The opening paragraph also potentially implied that the poll was more representative of scientists in the UK and greater EU than perhaps was warranted.

POCKET PSYCHIATRY



MOBILE MENTAL-HEALTH APPS HAVE EXPLODED ONTO THE MARKET, BUT FEW HAVE BEEN THOROUGHLY TESTED.

BY EMILY ANTHES

Type 'depression' into the Apple App Store and a list of at least a hundred programs will pop up on the screen. There are apps that diagnose depression (Depression Test), track moods (Optimism) and help people to "think more positive" (Affirmations!). There's Depression Cure Hypnosis ("The #1 Depression Cure Hypnosis App in the App Store"), Gratitude Journal ("the easiest and most effective way to rewire your brain in just five minutes a day"), and dozens more. And that's just for depression. There are apps pitched at people struggling with anxiety, schizophrenia, post-traumatic stress

disorder (PTSD), eating disorders and addiction.

This burgeoning industry may meet an important need. Estimates suggest that about 29% of people will experience a mental disorder in their lifetime¹. Data from the World Health Organization (WHO) show that many of those people — up to 55% in developed countries and 85% in developing ones — are not getting the treatment they need. Mobile health apps could help to fill the gap (see ‘Mobilizing mental health’). Given the ubiquity of smartphones, apps might serve as a digital life-line — particularly in rural and low-income regions — putting a portable therapist in every pocket. “We can now reach people that up until recently were completely unreachable to us,” says Dror Ben-Zeev, who directs the mHealth for Mental Health Program at the Dartmouth Psychiatric Research Center in Lebanon, New Hampshire.

Public-health organizations have been buying into the concept. In its Mental Health Action Plan 2013–2020, the WHO recommended “the promotion of self-care, for instance, through the use of electronic and mobile health technologies.” And the UK National Health Service (NHS) website NHS Choices carries a short list of online mental-health resources, including a few apps, that it has formally endorsed.

But the technology is moving a lot faster than the science. Although there is some evidence that empirically based, well-designed mental-health apps can improve outcomes for patients, the vast majority remain unstudied. They may or may not be effective, and some may even be harmful. Scientists and health officials are now beginning to investigate their potential benefits and pitfalls more thoroughly, but there is still a lot left to learn and little guidance for consumers.

“If you type in ‘depression’, its hard to know if the apps that you get back are high quality, if they work, if they’re even safe to use,” says John Torous, a psychiatrist at Harvard Medical School in Boston, Massachusetts, who chairs the American Psychiatric Association’s Smartphone App Evaluation Task Force. “Right now it almost feels like the Wild West of health care.”

APP HAPPY

Electronic interventions are not new to psychology; there is robust literature showing that Internet-based cognitive behavioural therapy (CBT), a therapeutic approach that aims to change problematic thoughts and behaviours, can be effective for treating conditions such as depression, anxiety and eating disorders. But many of these online therapeutic programmes are designed to be completed in lengthy sessions in front of a conventional computer screen.

Smartphone apps, on the other hand, can be used on the go. “It’s a way of people getting access to treatment that’s flexible and fits in with their lifestyle and also deals with the issues around stigma — if people are not quite ready to maybe go and see their doctor, then it might be a first step to seeking help,” says Jen Martin, the programme manager at MindTech, a national centre funded by the United Kingdom’s National Institute for Health Research and devoted to developing and testing new mental-health technologies.

One of the best-known publicly available apps was devised to meet that desire for flexibility. In 2010, US government psychologists conducting focus groups with military veterans who had PTSD learned that they wanted a tool they could use whenever their symptoms flared up. “They wanted something that they could use in the moment when the distress was rising — so when they were in line at the supermarket,” says Eric Kuhn, a clinical psychologist and the mobile apps lead at the US Department of Veterans Affairs’ National Center for PTSD.

The department joined up with the US Department of Defense to create PTSD Coach, a free smartphone app released in early 2011. Anyone who has experienced trauma can use the app to

learn more about PTSD, track symptoms and set up a support network of friends and family members. The app also provides strategies for coping with overwhelming emotions; it might suggest that users distract themselves by finding a funny video on YouTube or lead users through visualization exercises.

In its first three years in app stores, PTSD Coach was downloaded more than 150,000 times in 86 different countries. It has shown promise in several small studies; in a 2014 study of 45 veterans, more than 80% reported that the app helped them to

“IF YOU TYPE IN ‘DEPRESSION’, ITS HARD TO KNOW IF THE APPS THAT YOU GET BACK ARE HIGH QUALITY, IF THEY WORK, IF THEY’RE EVEN SAFE TO USE.”

track and manage their symptoms and provided practical solutions to their problems². More results are expected soon. Kuhn and his colleagues recently completed a 120-person randomized trial of the app, and a Dutch team is currently analysing data from a 1,300-patient trial on a similar app called SUPPORT Coach.

Smartphone apps can also interact with users proactively, pinging them to ask about their moods, thoughts and overall well-being. Ben-Zeev created one called FOCUS, which is geared towards patients with schizophrenia. Several times a day, the app prompts users to answer questions such as “How well did you sleep last night?” or “How has your mood been today?” If users report that they slept poorly, or have been feeling anxious, the app will suggest strategies for tackling that problem, such as limiting caffeine intake or doing some deep-breathing exercises.

Some apps help people to stay connected to health-care professionals, too. ClinTouch, a psychiatric-symptom-assessment app designed by researchers at the University of Manchester, UK, analyses users’ responses for signs that they may be experiencing a relapse; it can even notify a clinical-care team.

Small feasibility studies — which are generally designed to determine whether an intervention is practical, but do not necessarily evaluate its efficacy — have shown that patients use and like both apps, and a 2014 study found that those who used FOCUS for a month experienced a reduction in psychotic symptoms and depression³. FOCUS and ClinTouch are both now being evaluated in randomized, controlled trials.

Some researchers see opportunities in the data that smartphones collect about their users’ movement patterns or communication activity, which could provide a potential window into mental health. “Your smartphone is really this interesting diary of your life,” says Anmol Madan, the co-founder and chief executive of Ginger.io, a digital mental-health company based in San Francisco, California. Studies have now shown that certain patterns of smartphone use can predict changes in mental-health symptoms⁴; a drop in the frequency of outgoing text messages, for instance, may suggest that a user’s depression is worsening.

The Ginger.io app, which is still in beta, monitors these sorts of patterns and alerts each user’s assigned mental-health coach if it detects a worrying change.

ABSENT EVIDENCE

The evidence supporting the use of such apps is building^{5–7}. But this is a science in its infancy. Much of the research has been limited to pilot studies, and randomized trials tend to be small and unreplicated. Many studies have been conducted by the apps’ own developers, rather than by independent researchers. Placebo-controlled

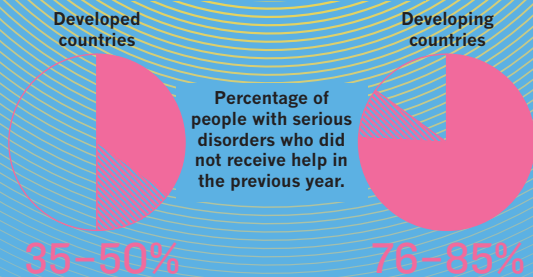
MOBILIZING MENTAL HEALTH

SMARTPHONE APPS FOR MENTAL HEALTH HAVE THE POTENTIAL TO REACH PEOPLE WITHOUT ACCESS TO CARE.

Global prevalence of mental disorders

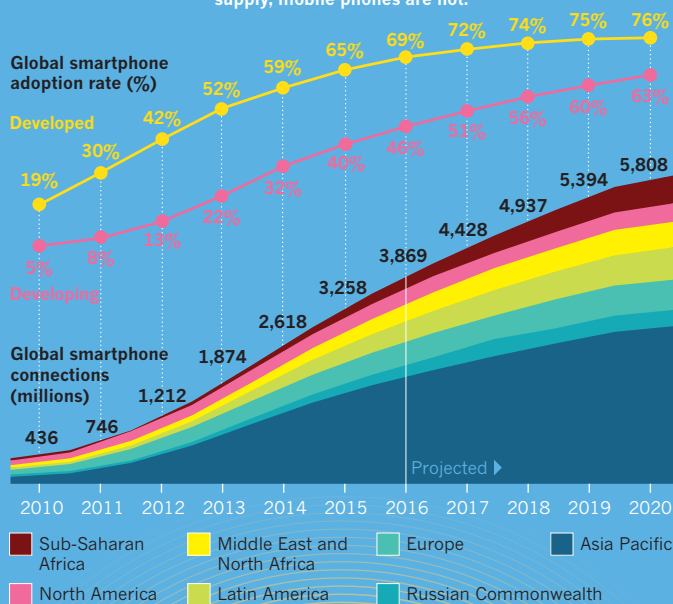
29%

Many people with mental illnesses don't get the help they need.



Although many factors might explain this treatment gap, a shortage of trained mental-health professionals plays a part, particularly in low-income countries.

But although psychiatrists may be in short supply, mobile phones are not.



Of about 15,000 disease-specific mobile health apps identified in a 2015 survey, nearly one-third dealt with mental-health issues.

29%

Disease-specific health apps that focus on mental health

trials are rare, raising the possibility that a 'digital placebo effect' may explain some of the positive outcomes that researchers have documented, says Torous. "We know that people have very strong relationships with their smartphones," and receiving messages and advice through a familiar, personal device may be enough to make some people feel better, he explains.

But the bare fact is that most apps haven't been tested at all. A 2013 review⁸ identified more than 1,500 depression-related apps in commercial app stores but just 32 published research papers on the subject. In another study published that year⁹, Australian researchers applied even more stringent criteria, searching the scientific literature for papers that assessed how commercially available apps affected mental-health symptoms or disorders. They found eight papers on five different apps.

The same year, the NHS launched a library of "safe and trusted" health apps that included 14 devoted to treating depression or anxiety. But when two researchers took a close look at these apps last year, they found that only 4 of the 14 provided any evidence to support their claims¹⁰. Simon Leigh, a health economist at Life-code Solutions in Liverpool, UK, who conducted the analysis, says he wasn't shocked by the finding because efficacy research is costly and may mean that app developers have less to spend on marketing their products.

A separate analysis¹¹ found that 35 of the mobile health apps originally listed by the NHS transmitted identifying information — such as e-mail addresses, names and birthdates — about users over the Internet, and two-thirds of these did not encrypt the data.

Last year, the NHS took this apps library offline and posted a smaller collection of recommended online mental-health services. The NHS did not respond to e-mailed questions or make an official available for interview, but it did provide this statement: "We are working to upgrade the Health Apps Library, which was launched as a pilot site in 2013 to review and recommend apps against a defined set of criteria which included data protection."

The regulation of mental-health apps is opaque. Some apps designed to be used in a medical context can be considered medical devices and therefore may be regulated by the UK Medicines and Healthcare Products Regulatory Agency, the US Food and Drug Administration (FDA) or equivalent bodies elsewhere. But the lines are fuzzy. In general, an app that claims to prevent, diagnose or treat a specific disease is likely to be considered a medical device and to attract regulatory scrutiny, whereas one that promises to 'boost mood' or provide 'coaching' might not. The FDA has said that it will regulate only those health apps that present the highest risks to patients if they work improperly; even mental-health apps that qualify as medical devices might not be regulated if the agency deems them to be relatively low risk.

But the potential risks are not well understood. "At the low end, people might waste their money or waste their time," says Martin, "and at the higher end, especially with mental health, they might be actively harmful or giving dangerous advice or preventing people from going and getting proper treatment."

When a team of Australian researchers reviewed 82 commercially available smartphone apps for people with bipolar disorder¹², they found that some presented information that was "critically wrong". One, called iBipolar, advised people in the middle of a manic episode to drink hard liquor to help them to sleep, and another, called What is Biopolar Disorder, suggested that bipolar disorder could be contagious. Neither app seems to be available any more.

Martin says that in Europe, at least, apps tend to come in two varieties, those that are commercially developed and come with little supporting evidence or plans for evaluation, and those with academic or government backing that take a more rigorous

SOURCES: WHO; REF. 1: GSMA; IMS INST.

approach. The problem is that the former are generally more engaging for users and the latter take so long to make it to the market — if they even do — that they look out of date. “This is a generalization,” Martin says, “but it’s broadly true.”

UNINTENDED CONSEQUENCES

Even well-intentioned apps can produce unpredictable outcomes. Take Promillekoll, a smartphone app created by Sweden’s government-owned liquor retailer, designed to help curb risky drinking. While out at a pub or a party, users enter each drink they consume and the app spits out an approximate blood-alcohol concentration.

When Swedish researchers tested the app on college students, they found that men who were randomly assigned to use the app ended up drinking more frequently than before, although their total alcohol consumption did not increase. “We can only speculate that app users may have felt more confident that they could rely on the app to reduce negative effects of drinking and therefore felt able to drink more often,” the researchers wrote in their 2014 paper¹³.

It’s also possible, the scientists say, that the app spurred male students to turn drinking into a game. “I think that these apps are kind of playthings,” says Anne Berman, a clinical psychologist at the Karolinska Institute in Stockholm and one of the study’s authors. There are other risks too. In early trials of ClinTouch, researchers found that the symptom-monitoring app actually exacerbated symptoms for a small number of patients with psychotic disorders, says John Ainsworth at the University of Manchester, who helped to develop the app. “We need to very carefully manage the initial phases of somebody using this kind of technology and make sure they’re well monitored,” he says.

In a pilot trial published earlier this year, ten US veterans with PTSD were randomly assigned to use PTSD Coach on their own for eight weeks, while another ten used the app with the support and guidance of primary-care providers. At the end of the trial, seven of the ten patients using the app with support showed a reduction in PTSD symptoms, compared with just three of the patients who used the app on their own¹⁴.

But if apps require medical supervision, that undermines the idea that they will serve as an easy and low-cost way to provide care to the masses. “People think there’s an app for everything,” says Helen Christensen, the director of the Black Dog Institute at the University of New South Wales in Sydney, Australia, who has developed and studied mental-health apps. “It’s actually about how we build systems around apps, so that people have health care.”

Distributing mental-health apps in the developing world presents further challenges. Although mobile technology is spreading rapidly, there are many people who do not have — or cannot afford — smartphones or mobile Internet access. And the content of apps needs to be delivered in local languages and reflect local cultures. “The notion that you can take an intervention and just plop it down in a region where people might not even use the same terms for mental health as you’re using is a little unrealistic,” says Ben-Zeev. “What we might call ‘hearing voices’ in the United States might be something like ‘communicating with your elders’ in a different region, depending on what label people attach to that experience.”

At this point, the notion that apps can deliver quality health care in low-income regions remains largely theoretical. “This is generally where the mHealth field is,” says Natalie Leon, a scientist at the South African Medical Research Council in Cape Town. “It’s a promise of potential effectiveness.”

GOOD PRACTICE

To make good on that promise, apps will have to be tested. Between 2013 and 2015, the number of mobile-health trials registered on ClinicalTrials.gov more than doubled, from

135 to 300. And the number of trials specifically focused on mental and behavioural health increased by 32%, according to a report by the IMS Institute for Health Informatics in Parsippany, New Jersey.

One digital health company that has earned praise from experts is Big Health, co-founded by Colin Espie, a sleep scientist at the University of Oxford, UK, and entrepreneur Peter Hames. The London-based company’s first product is Sleepio, a digital treatment for insomnia that can be accessed online or as a smartphone app. The app teaches users a variety of evidence-based strategies for tackling insomnia, including

“WE NEED TO VERY CAREFULLY MANAGE THE INITIAL PHASES OF SOMEBODY USING THIS TECHNOLOGY AND MAKE SURE THEY’RE WELL MONITORED.”

techniques for managing anxious and intrusive thoughts, boosting relaxation, and establishing a sleep-friendly environment and routine.

Before putting Sleepio to the test, Espie insisted on creating a placebo version of the app, which had the same look and feel as the real app, but led users through a set of sham visualization exercises with no known clinical benefits. In a randomized trial, published in 2012, Espie and his colleagues found that insomniacs using Sleepio reported greater gains in sleep efficiency — the percentage of time someone is asleep, out of the total time he or she spends in bed — and slightly larger improvements in daytime functioning than those using the placebo app¹⁵. In a follow-up 2014 paper¹⁶, they reported that Sleepio also reduced the racing, intrusive thoughts that can often interfere with sleep.

The Sleepio team is currently recruiting participants for a large, international trial and has provided vouchers for the app to several groups of independent researchers so that patients who enrol in their studies can access Sleepio for free.

“We think this is the way forward for digital health,” says Espie. Mobile-phone-based treatments, he says, “should be tested and judged like any other intervention. We shouldn’t treat people’s health with any less respect because the treatment is coming through an app.” ■ [SEE COMMENT P. 25](#)

Emily Anthes is a freelance journalist in New York City.

1. Steel, Z. *et al.* *Int. J. Epidemiol.* **43**, 476–493 (2014).
2. Kuhn, E. *et al.* *Mil. Med.* **179**, 12–18 (2014).
3. Ben-Zeev, D. *et al.* *Schizophr. Bull.* **40**, 1244–1253 (2014).
4. Beiwinkel, T. *et al.* *J. Med. Internet Res.* **3**, e23 (2016).
5. Dagö, J. *et al.* *J. Anxiety Disord.* **28**, 410–417 (2014).
6. Watts, S. *et al.* *BMC Psychiatry* **13**, 49 (2013).
7. Birney, A. J., Gunn, R., Russell, J. K. & Ary, D. V. *J. Med. Internet Res.* **4**, e8 (2016).
8. Martínez-Pérez, B., de la Torre-Díez, I. & López-Coronado, M. *J. Med. Internet Res.* **15**, e120 (2013).
9. Donker, T. *et al.* *J. Med. Internet Res.* **15**, e247 (2013).
10. Leigh, S. & Flatt, S. *Evid. Based Ment. Health* **18**, 97–99 (2015).
11. Huckvale, K., Prieto, J. T., Tilney, M., Benghozi, P.-J. & Car, J. *BMC Med.* **13**, 214 (2015).
12. Nicholas, J., Larsen, M. E., Proudfoot, J. & Christensen, H. *J. Med. Internet Res.* **17**, e198 (2015).
13. Gajecski, M., Berman, A. H., Sinadinovic, K., Rosendahl, I. & Andersson, C. *Addict. Sci. Clin. Pract.* **9**, 11 (2014).
14. Possemato, K. *et al.* *Gen. Hosp. Psychiatry* **38**, 94–98 (2016).
15. Espie, C. A. *et al.* *Sleep* **35**, 769–781 (2012).
16. Espie, C. A. *et al.* *Sleep Med.* **15**, 913–917 (2014).

SPRING BOOKS

PHYSICS

Soundtrack of the Universe

Sheila Rowan ponders a chronicle of the long road to gravitational-wave detection.

Gravitational waves were first detected on 14 September 2015 — a landmark event predicted a century before by Albert Einstein as part of his general theory of relativity. Janna Levin opens her chronicle of the long and twisting journey to the moment of detection, *Black Hole Blues and Other Songs from Outer Space*, with a quote from Niccolò Machiavelli's *The Prince*: "There is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things." Thus, words from a political treatise written in 1513 encapsulate the difficult birth of a new field: gravitational-wave astronomy.

The songs that Levin describes are the 'chirp' of almost unfathomably faint gravitational-wave signals: the soundtrack to



Black Hole Blues and Other Songs from Outer Space
JANNA LEVIN
Alfred Knopf: 2016.

our view of the skies. A signal arising from a collision between two black holes 400 million parsecs away was picked up by the US-led Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO). Levin, a theoretical cosmologist, tells the story of LIGO through interviews with key protagonists, including physicists Rainer Weiss, Ronald Drever, Kip Thorne and Rochus Vogt. She explains in clear terms the scientific heart of this achievement and the deep and personal fascination

that pursuing it has held for several generations of scientists. She also captures the cost of getting to this point, both financial — this is big science in its truest sense — and, in many cases, personal. In different ways, we are led to see how, from an idea, a calculation, a basic experiment in a laboratory, the scientists involved drove the field forward from the 1960s, individually and then together.

This is not the first book on LIGO's history. Sociologist of science Harry Collins, for instance, has written several, including *Gravity's Ghost and Big Dog* (University of Chicago Press, 2013; see M. de Laet *Nature* **501**, 164–165; 2013). Nor is it likely to be the last: the plot is too compelling. Verifying Einstein's prediction demanded precision technology and engineering beyond



ILLUSTRATIONS BY REBEKKA DUNLAP

anything achieved before. Optical instruments kilometres long were created, capable of sensing changes in the position of mirrors (induced by gravitational waves) of less than one ten-thousandth the diameter of a proton. In tandem arose a deep understanding of the astrophysical sources far out in the Universe that might be — and we now know are — sending gravitational signals detectable on Earth.

The work was, moreover, a tug of war between funders and colleagues with faith in the endeavour, and other parties who, for various reasons, felt that it did not warrant support. Levin describes this dissonance well; it is genuinely painful to read in places.

Levin is illuminating on how the very different Drever, Weiss and Thorne came to take the detection of gravitational waves as their life's work. She shows how individual efforts at the Massachusetts Institute of Technology in Cambridge (led by Weiss) and the California Institute of Technology in Pasadena (led by Drever and Thorne) came together in the 1980s through the realization that the detection of gravitational waves would be possible only through a combined

effort. In fact, they prefaced their 1989 funding proposal to the US National Science Foundation with the quote from *The Prince* that Levin uses. They were aware even then

SCIENTISTS BRING THEIR WHOLE HISTORY TO THEIR WORK: WHO THEY ARE SHAPES WHAT THEY DO.

of the difficult road ahead, although not the specific obstacles that would litter it.

For instance, as Levin shows, within a few years the project's leadership changed

in unforeseen ways. Weiss and Drever, both strong leaders with individual ideas on process, and both passionate about the science, could not agree on a shared vision — even with Thorne's mediation. Decision-making became impossible, threatening to derail the project. Vogt was appointed to restore some measure of stability.

Scientists bring their whole history to their work: who they are shapes what they do, how they behave as leaders and how they make the research happen. So it is salutary to read about Drever's background and its part in his ingenuity and thrift. He produced a stream of remarkable scientific ideas, sometimes starting with limited means and materials. Levin describes Drever in Glasgow, UK, carrying out research on Lorentz invariance — testing fundamental aspects of the theory of relativity using basic materials in his back garden. But such inherent creativity does not necessarily sit easily with project management or the shared responsibilities that are part of leading a large scientific team.

Similarly, Levin describes how Vogt's management style was shaped by a dislike of authority that stemmed from experiences in his native Germany in the 1940s. This, too, became problematic in the stewardship of LIGO, which, among other qualities, demanded the ability to take the strain of significant external oversight and formality. Further changes in directorship followed, bringing the LIGO project to the stable, successful state in which it remains today. Levin touches on those only lightly.

Despite the upheavals, 'unreasonableness' — that singular drive on which "all progress depends", according to playwright George Bernard Shaw — was clearly essential to the forging of this new scientific field. Weiss, Drever and Vogt all had that crucial, extraordinary ability to persevere against prevailing forces to enable progress.

In many passages in *Black Hole Blues*, it is clear that we are reading the result of splitting the white light of truth through the prism of memory. In focusing mostly on the captivating details of a few individuals, the full picture is lost, the whole remains largely unsynthesized, and gaps at the joins sometimes show. Most noticeably, the details in Levin's chronicle stop around 20 years ago. We skip from the end of Vogt's tenure as LIGO director in the mid-1990s to the point at which the advanced LIGO instruments are about to take data in 2015. As interesting as Levin's narrative is, it leaves an entire exciting book still to be written. ■

Sheila Rowan is director of the Institute for Gravitational Research at the University of Glasgow, UK.
e-mail: sheila.rowan@glasgow.ac.uk



METEOROLOGY

The brainstormers

Alan Thorpe enjoys a hymn to some of the founders of the science and institutions of weather forecasting.

It is thanks to the efforts of an international community of meteorologists and atmospheric scientists that accurate forecasts of major weather systems can be made reliably up to about a week ahead (see P. Bauer *et al.* *Nature* **525**, 47–55; 2015). Many researchers contributed to the revolution in weather sciences in the first half of the twentieth century, so it is perhaps invidious to single out a few.

Science historian James Fleming focuses on three: Norwegian Vilhelm Bjerknes, Swede Carl-Gustaf Rossby and American Harry Wexler. The first two I expected; the third I was intrigued to learn more about. Fleming devotes about 60 pages to each man's life and work, and mentions many others and their

impacts, mostly on US weather forecasting.

To the cognoscenti, the essentials of Bjerknes's and Rossby's science will be familiar. Fleming's fascinating account clarifies why these two were giants of leadership. Bjerknes created the 'Bergen school' of meteorology, which used rigorous scientific principles to understand and predict the evolution of weather features such as fronts and cyclones. The school included his son Jacob among many talented, mainly Scandinavian, scientists. Rossby established university schools of meteorology in Stockholm and Chicago, and at the Massachusetts Institute of Technology in Cambridge. The breadth of Wexler's role emerges through his contributions to the development of techniques and operational

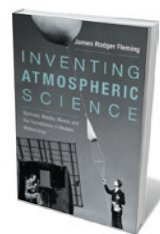
weather forecasting, particularly in the United States.

Between the birth of Bjerknes — the oldest — in 1862 and the death of Wexler, the youngest, in 1962, there passed a formative and innovative century. As Fleming reveals, their lives were linked, with Bjerknes teaching Rossby and Rossby, Wexler.

In 1904, in 'Weather forecasting as a problem in mechanics and physics', Bjerknes set the agenda for applying the laws of physics to the atmosphere to predict the weather (V. Bjerknes *Meteorol. Z.* **21**, 1–7; 1904). His vision was to use a sufficiently accurate knowledge of the state of the atmosphere and the laws that govern its evolution to forewarn people about weather to come. His motivation was to make his mark in what was for him a new field of science — he began his career working with his father, a physicist at the University of Oslo, on fluid analogies for the electromagnetic field. He was eager, too, to provide practical advice on hazards that affected mariners, farmers and the public.

Fleming notes the absence of a book-length biography of Rossby, and I hope that this will be rectified soon. To me, he is a first among equals. As well as building institutions, he established important principles, such as the conservation of potential vorticity — used to understand the development of rotation in cyclones and other weather systems — and the large-scale atmospheric wave patterns named after him. He was a leader in developing techniques such as experiments that simulate the atmosphere in a rotating tank of water, as well as aircraft soundings and the use of radiosondes, or radio-based measurements using weather balloons. A polymath with a high public profile, he was pictured on the cover of *Time* magazine in December 1956, with the title "Weatherman".

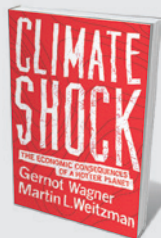
Wexler's contributions include making the first research flight into a hurricane, using radar to track storm systems, working



Inventing Atmospheric Science: Bjerknes, Rossby, Wexler, and the Foundations of Modern Meteorology
JAMES RODGER FLEMING
MIT Press: 2016.

NEW IN
PAPERBACK

Highlights of this
season's releases



Climate Shock: The Economic Consequences of a Hotter Planet

Gernot Wagner and Martin L. Weitzman (Princeton University Press, 2016)

Global, long-term, irreversible, uncertain: four words used by economists Gernot Wagner and Martin Weitzman in their timely warning on the impacts of climate change. It is not enough to simply hope that we are wrong about worst-case scenarios; the authors' most extreme predictions include a 20-metre sea-level rise and average global temperatures reaching 6°C above pre-industrial levels. Wagner and Weitzman urge us to act now to insure Earth against uncertainty.

towards space observations and developing the use of computers in meteorology.

From the findings of these three men, Fleming expertly weaves a tapestry of broader developments, from early uses of computers and satellites to numerical predictions with supercomputers. He explores intentional weather modification, radioactive fallout, rocketry, air pollution and electromagnetism. For example, the air movements made apparent when researchers tracked the fallout from nuclear-bomb tests in the 1950s provided insight into atmospheric circulation.

The penultimate chapter covers what Fleming calls the birth of atmospheric science in the late 1950s, coinciding with planning for the International Geophysical Year in 1957–58. In 1956, Rossby proposed enlarging the definition of meteorology to include elements such as atmospheric chemistry and relevant biological processes. Much of what we regard as contemporary developments, such as the understanding of atmospheric composition or geoengineering, were in the minds of 1950s researchers, Fleming points out.

Modifying the reflectivity of the planet to avoid harmful climate change was discussed by Wexler and others as early as 1958. In a 1962 speech, Wexler said: “We are in weather control now whether we know it or not.” Fleming also focuses on the importance of committees in planning developments such as the creation of the US National Center for Atmospheric Research in 1960, and the role of top researchers in leading these committees.

A historical account has to have boundaries, and Fleming barely hints at what came after 1960. I found the discussion of Edward Lorenz’s work on chaos that led to aspects of probabilistic forecasting disproportionately brief.

What shines through *Inventing Atmospheric Science* is the commitment of three men to applications of research to society, and their desire to advance our understanding of weather. This is an inspirational story, very well told. ■

Alan Thorpe is visiting professor of meteorology at the University of Reading, UK. e-mail: alan.thorpe@gmx.com

ENERGY

Oilman at the peak

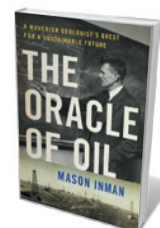
Gregor Macdonald applauds a biography of prescient geologist and energy theorist Marion King Hubbert.

A scientist’s work does not always intersect neatly with the events of their time, but that was the good fortune of US geologist and oilman Marion King Hubbert (1903–89). After labouring for decades to perfect forecasting of the oil-production limit, he saw his efforts validated in the energy crises of the early 1970s. His approach would later be known as Hubbert peak theory.

Journalist Mason Inman’s meticulous biography *The Oracle of Oil* follows Hubbert from his youth on the plains of Texas through the Great Depression, the Second World War and the rise of US President Ronald Reagan in 1981. But its scope is much more expansive. In Hubbert’s story, Inman has found a meditation on the booms and busts that marked twentieth-century economic growth. Hubbert’s iconoclastic career forms a perfect arc, from oil’s troubling oversupply in the 1920s to its relative scarcity after the peak of US production in 1971, when the US economy suffered oil shocks.

As Inman shows, Hubbert’s impact extends beyond oil: it is an early manifestation of ecological economics. At the end of his career, Hubbert remained concerned about nuclear waste; was convinced that high rates of growth were environmentally destructive; and conjectured that solar power might be the most viable energy solution. With regard to growth or sustainability, Hubbert’s work is an overlooked contribution to US economic history.

Arriving at the University of Chicago in Illinois in 1924 with little money but no



The Oracle of Oil: A Maverick Geologist’s Quest for a Sustainable Future
MASON INMAN
W. W. Norton: 2016.

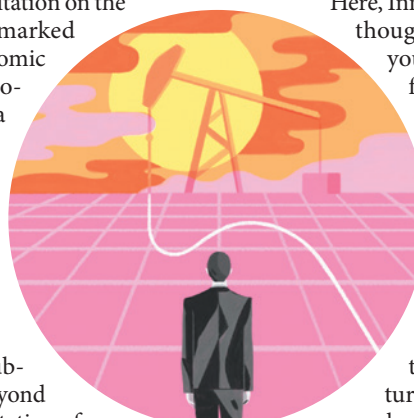
shortage of ambition, Hubbert saw geology as a wide-open field, and gate-crashed it. With precocious brilliance, he began to identify gaps and disorganization in its practices, and published his first paper, on fault classifications, as an undergraduate. In 1930, he was head-hunted by Columbia University in New York City to direct a new effort in geophysics.

Here, Inman’s biography reveals its thoughtful design, tracing the youthful roots of Hubbert’s formidable achievements.

The undergrad pondering Earth’s folds one summer later becomes a breakthrough geologist, solving complex scaling problems. The boy surrounded by oil rigs becomes a Shell executive, aggressively pursuing the hydrological and structural complexities of oilfield exploration. The young man who

questioned religious faith becomes a nonconformist US Geological Survey (USGS) researcher, insisting that popular forecasts are built less on data than on optimism.

Inman’s direct, explanatory style is well suited to describing the evolution of Hubbert’s thinking. At Columbia, Hubbert began to ponder the S-curve of growth that has long fascinated observers of economies, biology and natural-resource extraction. He perfected his technique over decades, from rough ▶



The Upright Thinkers

Leonard Mlodinow (Vintage, 2016)

Carrier pigeons once toted stock prices. Today, instant messaging manages the job. Theoretical physicist Leonard Mlodinow explores how the most human of desires, a thirst for knowledge, grew from Neanderthal hunger pangs to measuring our planet’s orbit around the Sun.



Why Information Grows

César A. Hidalgo (Penguin, 2016)

Economies are computers and information is at war with entropy, claims statistical physicist César Hidalgo. He shows how the scientific imagination needs knowledge and resources to grow, such as the Chilean copper that ‘feeds’ electronics (see Philip Ball’s review: *Nature* **521**, 420–421; 2015).

► estimates of ultimately recoverable US oil reserves to his eventual winning model: an advanced calculation that incorporated past production, yield per foot of exploration and the tricky variable of reserve growth.

Inman does not, however, cite the work of UK economist William Stanley Jevons, whose 1865 warning about the economy's over-reliance on coal prefigures the Hubbert story. Jevons died in 1882, so never saw his prediction come true: British coal output peaked in 1913. Hubbert, by contrast, was feted with numerous awards, including the Rockefeller Service Award in 1977, and broad coverage in *The New York Times* when his previous reports for the government and the USGS were acknowledged for their accuracy.

Hubbert's forecast was not the end of the US oil story. After his death, production continued to languish, in accordance with his forecast. But with fracking, the United States lifted oil production as recently as last year to levels close to the 1970 peak. Oil production is now falling again owing to a price bust — global supply capabilities were created for demand that failed to materialize. Inman does a fine job of handling this recent history.

The Oracle of Oil offers valuable insights beyond energy. In the demand-side bust of the 1930s, it shows Hubbert thinking deeply about the surplus of labour created partly, in his view, by the effects of powerful oil married to the newest machines: cars, construction equipment and aircraft. Hubbert was co-founder of Technocracy, a group of New York intellectuals aiming to prevent future economic dislocations. Two publications by keen observers of the low-growth problem — Thomas Piketty's *Capital in the Twenty-First Century* (Seuil, 2013) and Robert Gordon's 2012 paper 'Is US Economic Growth Over?' (see go.nature.com/wblxig) — also explore this territory of limits and sustainability.

In Inman's work, the oilman emerges as a restless and prescient figure concerned with the environment. In writing the first biography of Hubbert, Inman has retrieved, if not rescued, the story of a scientist who has much to offer to today's energy conundrum. ■

Gregor Macdonald is a journalist based in Portland, Oregon, who covers the energy sector.

e-mail: gregor@gregor.us

NEUROSCIENCE

Listening in on yourself

Douwe Draaisma is intrigued by a study examining both 'the voice within' and verbal auditory hallucinations.

Thinking about thinking is a curious exercise. Most of us probably agree that much of our own thought process takes the form of inner speech. But would we also agree that we hear an inner voice? If not, why would we call it talking at all? Can we experience and observe inner speech simultaneously, or would this be like "trying to turn up the gas quickly enough to see how the darkness looks", as US psychologist William James asked himself in 1890? Each of these questions may lead one into a philosophical forest, dense and dark.

Side-stepping such conceptual intricacies, psychologist Charles Fernyhough convincingly explores inner speech from a practical perspective. In *The Voices Within*, he discusses how people with aphasia (a speech and language disorder that stems from brain damage) may lose their sense of inner speech; how deaf people 'talk to themselves' (mostly in sign language, some by lipreading); how more than 60% of children have had silent conversations with imaginary friends; and whether people who stutter experience their inner speech as fluent (they do). Silent self-talk, evasive as it is to introspection, turns out to be a robust and quintessential part of memory, thought and imagination.

Fernyhough's sources are equally wide-ranging. He draws on internal monologues



The Voices Within: The History and Science of How We Talk to Ourselves
CHARLES FERNYHOUGH
Profile: 2016.

in Gustave Flaubert's 1856 novel *Madame Bovary* and reports of self-talk by professional cricketers. He mentions physicist Richard Feynman having an argument with himself, and Joan of Arc insisting that God talked to her in French, not Latin. Today, Fernyhough directs *Hearing the*

Voice, a research project at Durham University, UK, funded by biomedical charity the Wellcome Trust.

When I'm invited to write a review, I know from experience that it is wise to switch my inner speech from Dutch, my native language, to English, which I routinely use for scientific communi-

cation. Most bilingual people have no trouble identifying the language that they are thinking in. But I would be hard-pressed to say whether I talk to myself at a natural speed or in an abbreviated way, much less whether the stream of my thoughts flows equally fast (or slow) in both languages. Most people say that they have the definite impression that their inner speech unfolds faster than actual speech.

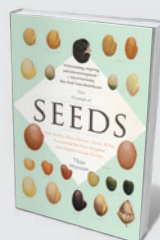
**SILENT SELF-TALK
TURNS OUT TO BE A
ROBUST
PART OF
MEMORY.**



Spirals in Time: The Secret Life and Curious Afterlife of Seashells

Helen Scales (Bloomsbury, 2016)

From beachcombing to shipwreck diving, marine biologist Helen Scales shares her love of molluscs, many of which convert seawater into protective homes. Shells, she reveals, have served as everything from jewellery to calcareous currency.



The Triumph of Seeds

Thor Hanson (Basic, 2016)

Biologist Thor Hanson sows the ultimate celebration of seeds and how they conquered Earth. Kernels can be crafty: unripe fruit, for instance, tastes bitter to deter predators from dispersing the seeds too soon (see Sandra Knapp's review: *Nature* **519**, 288–289; 2015).



Referring to the ideas of philosopher Eric Schwitzgebel, Fernyhough points out that this leaves several questions open. Is internal speech faster because it is unhampered by slower motor processes, or because we don't think in full sentences? According to twentieth-century Russian psychologist Lev Vygotsky, inner speech develops as children's conversations 'go underground', often as a result of social encouragement. In the process, they begin to think in 'pure meanings', causing telegraphic inner speech. This condensation, in turn, could help to explain the paradoxical sensation that inner speech feels faster, but never rushed. Sped up or condensed? To this day, there are no consensual methods or techniques to decide between the two.

Fernyhough also presents several interviews with participants in his Durham research project, itself an offshoot of the Hearing Voices Movement founded in 1987 by Dutch psychiatrist Marius Romme and now expanded to 23 countries. At present, the United Kingdom has 180 support groups for voice-hearers. Some people report hearing intrusive voices, often whispering or

shouting abuse, and seek relief by joining a local group. Most people who experience verbal auditory hallucinations have had a diagnosis of schizophrenia or have experienced childhood abuse — but, as Fernyhough points out, not all. Clinically relevant as they are, these chapters expose a tension in *The Voices Within*: it is still a matter of controversy whether hearing voices has much to do with the quiet self-talk of ordinary thinking.

Fernyhough tries to bridge this gap with a model that he dubs dialogic thinking, which conceptualizes inner speech as an internalized conversation between different voices. He hypothesizes that if a patient fails to identify a particular utterance as a fragment of some inner dispute, he might experience this fragment as coming from an external source — a hallucination. But many voice-hearers also experience inner speech, and can distinguish between the two. It remains to be seen whether the experience of hearing voices will really offer a window on inner speech.

We have come a long way from US psychologist John Watson's behaviourist speculation

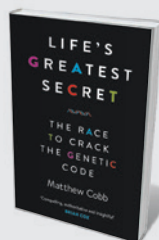
that inner speech is simply covert motor action in the speech apparatus. New methods have invited new distinctions. Neuroimaging studies by Fernyhough and his colleagues suggest that speaking internally when instructed activates Broca's area in the brain (associated with producing speech), and deactivates Heschl's gyrus (associated with auditory perception). Spontaneous, free-flowing inner speech, on the other hand, involved the opposite pattern of neural activation. This is but one of many promising avenues of research, and *The Voices Within* is full of them. Profound and eloquent (he is a novelist too), Fernyhough presents an intriguing array of fresh findings and perspectives. He makes a persuasive case that one of the most intimate and private of our mental activities has a social origin. We talk to ourselves because we talked to others first. ■

Douwe Draaisma is professor of the history of psychology at the University of Groningen in the Netherlands. His latest book is *Forgetting*.
e-mail: d.draaisma@rug.nl



Infested: How the Bed Bug Infiltrated Our Bedrooms and Took Over the World

Brooke Borel (University of Chicago Press, 2016)
Bed bugs (*Cimex lectularius*) are perfectly adapted for bloodsucking. Toothed mandibles pierce their victims' skin and inject saliva proteins that widen the blood vessels and prevent clotting, reveals Brooke Borel in her creepy exposé of the household pest.



Life's Greatest Secret: The Race to Crack the Genetic Code

Matthew Cobb (Profile, 2016)
Anecdotes abound in zoologist Matthew Cobb's history of the quest to unravel the genetic code. Cobb updates the story with a look at gene-editing tool CRISPR and its role in gene therapy, agriculture and the control of invasive species.

The women who launched NASA

Jennifer Light savours the history of a doughty band of ‘human computers’.

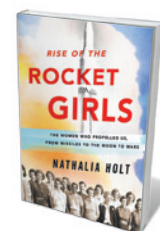
Data is having its moment in the sun. It used to be an incidental detail in stories about personalities and institutions in the history of science and engineering, politics and culture. Now, the construction of databases and details of data analysis are featured events. Even the 2015 film *Spotlight*, which dramatizes a journalistic investigation into sexual abuse in the Catholic Church, has an extended scene in which the protagonists compile a spreadsheet. Number crunchers, once stereotyped as missing the bigger picture, now are the big picture.

Natalia Holt's *Rise of the Rocket Girls* reveals how, from the 1940s, a group of ‘crunchers’ operated in near secrecy at NASA — their anonymity as much to do with their gender as with the status of data. These women staffed NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California, as human computers and, later, programmers. Most held degrees in mathematics or science. Their work, Holt reveals, was innovative, yet was considered less important than other JPL outputs. Holt documents the work of previously unknown figures such as Barbara Canright, Susan Finley and Helen Ling, who supplied crucial calculations that helped to launch missiles, bombers and the first US satellite; to control lunar missions; and even to navigate today's Mars rovers. Intercut with the human stories, Holt carefully lays out practical problems — such as the need to minimize fuel weight while ensuring that a rocket could reach escape velocity — and how mathematics helped to solved them.

Here, maths is dramatic, not mundane. Calculating is a physical, even athletic, act. Holt describes Canright's experience: “Her right index finger was lined with thick red and white calluses, the result of clutching a pencil for hours a day. Her grip on the pencil



often made her hand perspire, leaving pucker marks across the graph paper.” The women held “computing races” to solve complex mathematical problems in their down time. But as Holt shows, ‘doing the maths’ was often a matter of life and death. She describes Marie Crowley's work on the ideal size and shape of a rocket's nozzle opening: the faster a nozzle can clear exhaust, the more thrust the rocket generates. On one occasion, as engineers were about to run a test based on her calculations,

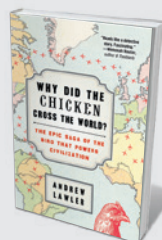


Rise of the Rocket Girls: The Women Who Propelled Us, from Missiles to the Moon to Mars
NATHALIA HOLT
Little, Brown: 2016.

Crowley realized that she had forgotten to take a square root. Her phone call came too late and the rocket failed — luckily, no one was hurt. Often rushed and never double-checked in those seat-of-the-pants days, the calculations were a source of anxiety for the women.

The drama of data is not the book's main rationale. Holt's larger interest is that it was women who did the calculating. The book is organized chronologically, with photographs of starring figures at the start of each section. In keeping with larger currents in historical studies of databases and data analysis, Holt argues that these women's calculations played an under-appreciated part in NASA's towering achievements. She bases her chronicle on interviews with several JPL data crunchers, members of their families and lab engineers, as well as a reading of scholarly and archival literature on the space programme.

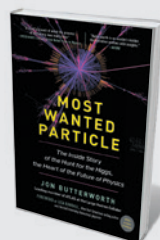
The book does not show equal mastery of research on gender and labour in technical fields: Holt provides more data than analysis. She depicts the human computers' life stories vividly, but her reflections on larger themes are limited. One is the relationship between people and machines. For example, Holt mentions the women's limited use of calculation aids. They often found it faster to do hand calculations than to work with a Friden calculating machine. Even in the late 1950s, when JPL acquired an IBM 704 — an early mass-produced computer — the crunchers (and engineers) were



Why Did the Chicken Cross the World?

Andrew Lawler (Atria, 2016)

With 20 billion chickens roaming Earth at any one time, this beleaguered bird truly merits Andrew Lawler's illuminating eulogy. It shows how the red junglefowl (*Gallus gallus*) strutting in a Borneo jungle became an economic mainstay (see Ewen Callaway's review: *Nature* **515**, 490–491; 2014).



Most Wanted Particle

Jon Butterworth (*The Experiment*, 2016)

Experimental physics becomes accessible as well as astounding in this insider's account of the hunt for the Higgs boson, detected in 2012. Explaining fermions, jet algorithms and particle-accelerator malfunctions with ease, Jon Butterworth puts a lively spin on atomic science.

wary of machines “that had too many glitches to be trustworthy”. The speed and accuracy of later models threatened to make the mathematicians’ roles redundant, but the women staved off obsolescence for some time by redefining their official jobs as programmers. Holt misses an opportunity here to make the connection between the experiences of JPL’s human computers in relation to the rise of automation, and today’s widespread anxieties over the ‘age of artificial intelligence’.

Personal anecdotes include details of the family arrangements that the women had to make in an era when marriage and children meant leaving the workforce. These will fascinate general readers and provide valuable primary source materials for future academics. Yet Holt does not contextualize the JPL crunchers’ experience in the broader history of women in science and technology. She does observe that the proportion of JPL technical jobs held by women — 15% in the 1990s — was partly due to women organizing the hiring process. Yet although the figure was higher than elsewhere in NASA, it was still remarkably low. And in the same way that NASA failed to celebrate these women’s contributions, it also overlooked the evidence in favour of including women in the astronaut corps. As space curator Margaret Weitekamp has detailed in *Right Stuff, Wrong Sex* (Johns Hopkins University Press, 2004), the benefits of sending women to space included their small size compared to men, which meant that they would need less food and water.

Engagement with such studies would have helped Holt to analyse how typical the JPL programmers’ experience was for women in technical fields, and to explicate her original findings more clearly. Many scholarly studies describe the exodus of women at other institutions from programming by the 1950s. By contrast, Holt notes, some of the early female crunchers stayed at JPL well into the twenty-first century, some in managerial roles. That raises questions about what made JPL different and, crucially, about retaining women in technical fields today. That story remains to be written. ■

Jennifer Light is professor of science, technology and society at the Massachusetts Institute of Technology in Cambridge.
e-mail: jslight@mit.edu

GENETICS

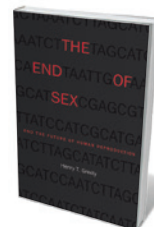
Coitus defunctus

Lori Andrews assesses Henry Greely’s treatise on how technology will oust reproductive intimacy.

For millennia, people have been trying everything from magic to medicine to influence the traits of their children. Recommendations from relatives and advice columns have, across time, included: put a knife under your bed if you want a boy; eat sweets while pregnant if you want a girl; place headphones on your pregnant belly and play Jacopo Peri’s *Euridice* to ensure that your future child will appreciate opera.

Current genetic technologies offer parents more-precise means of predicting and perhaps shaping the traits of their children. Embryos created through *in vitro* fertilization (IVF) can be analysed using pre-implantation genetic diagnosis (PGD); their entire genomes can even be sequenced. And the gene-editing technique CRISPR–Cas9 is expected to one day offer couples a chance to repair and enhance the genes of their embryos.

In *The End of Sex and the Future of Human Reproduction*, lawyer and bioethicist Henry Greely does an enviable job of explaining the scientific underpinnings and legal regulation of current reproductive and genetic technologies. The central focus of his book is his prediction that a new technology will develop — one that he dubs “Easy PGD”. Greely envisions a situation in which a woman will not have to undergo treatment with hormones and have her eggs removed to produce an embryo for testing — as is done in the course of IVF. Instead, some of her skin cells will be removed and coaxed by stem-cell technologies to turn



The End of Sex and the Future of Human Reproduction

HENRY T. GREELY
Harvard University Press:
2016.

into eggs. The eggs will then be fertilized by sperm from her partner to create as many as 100 embryos. In this scenario, parents will be able to choose which embryos to have implanted in the woman’s uterus on the basis of hundreds of traits (or more) revealed by whole-genome sequencing.

Greely makes three claims about this putative Easy PGD. First, it will replace sex as a way to create babies. Second, it will be

more socially acceptable than current PGD and prenatal genetic-testing technologies. And third, it will be free to the user.

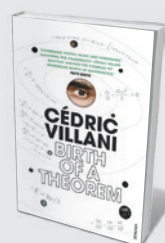
All three claims are problematic.

Greely assumes that people will elect to create children with Easy PGD rather than through sex because of the desire to control the traits of their chil-

dren. But if people were so keen to choose embryos on the basis of their genotypes, all couples who use IVF would submit their embryos to genetic testing (given that this would involve no extra risk to the woman). In the United States, only 5% of such couples do so.

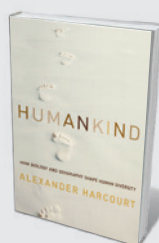
Greely also speculates that couples will rush to use stem cells to create embryos because such an approach would be less ▶

THE PROCESS OF DECISION- MAKING WOULD BE DAUNTING.



Birth of a Theorem: A Mathematical Adventure

Cédric Villani (Vintage, 2016)
Tackling the 140-year-old Boltzmann equation (with Clément Mouhot) led mathematician Cédric Villani to win a share in the 2010 Fields Medal. Documenting this quest, Villani encapsulates the despair and elation that maths can incite (see Amir Alexander’s review: *Nature* **519**, 31–32; 2015).



Humankind: How Biology and Geography Shape Human Diversity

Alexander Harcourt (Pegasus, 2016)
Biogeographer Alexander Harcourt ponders the myriad forces that led to the amazing diversity of *Homo sapiens* as we spread across the globe. He thinks that coastal migration once prevailed: harvesting seafood was easier than hunting.



► invasive than IVF. But there is no proof that the risks to women presented by IVF (such as ovarian hyperstimulation syndrome and pelvic infection) are a substantial deterrent when it comes to couples who use the procedure: globally, more than 400,000 children are born through IVF each year. Nor is there any evidence that couples would embrace an untested stem-cell procedure, with unknown risks to the resulting child.

Even if Easy PGD were risk-free, the process of decision-making would be daunting. Whole-genome sequencing would

provide prospective parents with hundreds or thousands of bits of information about each embryo. The *BRCA1* tumour-suppressor gene alone consists of more than 81,000 base pairs. Of the hundreds of *BRCA1* mutations that have been reported, some are associated with an increased risk of developing breast cancer or ovarian cancer; others are not, and the significance of many is not yet known.

If the test reports a mutation of unknown significance in any of an embryo's 20,000 genes, parents might end up discarding a perfectly healthy embryo, or — if

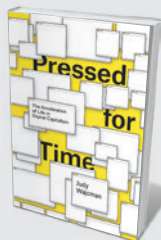
they have the embryo implanted anyway — worrying throughout the child's life about the potential manifestation of a horrible condition. Even the mutations that are associated with disease will be hard to make sense of. What does a person do with the knowledge that a certain embryo will develop into someone who has a breast-cancer risk 30% higher than that of the general population, double the baseline risk of developing Alzheimer's disease, and a 45% possibility of achieving a higher college admission score than another embryo with an equally complex data set?

Greely claims that Easy PGD will be more socially acceptable than current genetic testing of embryos and fetuses because, unlike existing approaches, it does not involve abortion. Yet whether the embryos are derived from skin cells or seeded with gametes, right-to-life advocates will oppose the termination of embryos that could be implanted to create children. Furthermore, many more embryos would be discarded using Greely's proposed Easy PGD than are now destroyed in the course of IVF and PGD.

Greely's final claim is that Easy PGD will be free. Using assumptions, including that it will cost US\$1,000 to create 100 embryos from skin cells, \$60 per embryo for genetic diagnosis, and \$500 for genetic counselling, he estimates that it will cost about \$11,000 for the entire procedure. According to Greely, at that cost, "the procedure should pay for itself for health reasons". But current genetic technologies also save future health costs, and these techniques are not always covered by health-care systems. In fact, it is extremely difficult for women below a certain income bracket to access them.

Greely imagines the slogans that clinics will use to tout Easy PGD: "You want the best for your child; why not have the best child you can?" But after exploring Greely's claims in detail, Easy PGD seems more likely to be an expensive technology of unknown risks that would present parents with hard-to-interpret information about as many as 100 embryos. The idea of Easy PGD should make us uneasy indeed. ■

Lori Andrews is distinguished professor of law at Chicago-Kent College of Law, Illinois Institute of Technology.
e-mail: landrews@kentlaw.iit.edu



Pressed for Time: The Acceleration of Life in Digital Capitalism

Judy Wajcman (University of Chicago Press, 2016)
Sociologist Judy Wajcman sagely analyses the disparate experience of time as technology has evolved. Despite the common belief that smartphones heighten stress, she argues that we are not victims of machines, but masters of their role in our lives.



The Cosmic Cocktail: Three Parts Dark Matter

Katherine Freese (Princeton University Press, 2016)
What is the Universe made of? Physicist Katherine Freese chronicles the cracking of this beguiling enigma, from the eccentric, ski-jumping Fritz Zwicky (who coined the term dark matter) to particle-smashing physics (see Francis Halzen's review: *Nature* **509**, 560–561; 2014). **Emily Banham**

Correspondence

Embed stormwater use in city planning

Potable water resources are being depleted at an alarming rate worldwide. Storm water is a hugely under-utilized resource that could help as extreme weather events become more frequent.

The challenges of collecting and using storm water mean that the practice is not widespread. Rainfall tends to be seasonal, so storm water must be stored for use in dry periods in natural underground aquifers (see A. Mankad *et al.* *J. Clean. Prod.* **89**, 214–223; 2015) or in specially built reservoirs. The reliance of such projects on the weather can make the costs hard to justify.

Storm water may also be heavily polluted and is expensive to treat. This can make alternatives such as imposed water rationing or water transfers from other areas more attractive — despite their human and environmental costs. Stormwater treatment would be more economically viable if less-purified water were used for non-drinking purposes. This would require wider public education, because the idea of recycled water is anathema to many.

For expanding cities, ways to deploy storm water will need to be embedded into urban planning. The technology is already available and could be tailored to different situations (see, for example, T. D. Fletcher *et al.* *J. Environ. Qual.* **37**, S-116–S-127; 2008; A. E. Barbosa *et al.* *Water Res.* **46**, 6787–6798; 2012).
Shunsuke Managi *Kyushu University, Fukuoka, Japan.*
Ashantha Goonetilleke, Clevo Wilson *Queensland University of Technology, Brisbane, Australia.*
managi@doc.kyushu-u.ac.jp

Engage engineers in soil management

The UK government's Environmental Audit Committee is due to release a report on its enquiry into soil health by this

summer. With most soil guidance for the United Kingdom coming from the European Union and the United Nations, the possibility of Britain leaving the EU ('Brexit') makes UK soil protection even more pressing.

US President Franklin D. Roosevelt declared in 1937 that "the nation that destroys its soils destroys itself" (see also L. Montanarella *Nature* **528**, 32–33; 2015). Since agriculture became industrialized, we have been taking more organic matter out of soils than we put in. Together with urban development, flooding and only partial return of organic wastes to soil, this means that soil carbon is being depleted globally, not just in the United Kingdom.

Engineers could lead the way in redressing this balance, as they have for recycling precious metals such as catalytic platinum (see, for example, C. Hagelüken *Platin. Met. Rev.* **56**, 29–35; 2012). For instance, they can stabilize carbon in organic wastes using minerals such as manganese oxides (K. Johnson *et al.* *Nature Commun.* **6**, 7628; 2015) and optimize soil amendments to maintain and enhance soil ecosystem services. Engineers, who are currently largely excluded from sustainable soil management, will then be working with the environment — instead of just on top of it.

Karen L. Johnson *Durham University, UK.*
karen.johnson@durham.ac.uk

A global baseline for ecosystem recovery

Indicators of success for ecosystem restoration depend on a proper assessment of the extent to which human activity has degraded biodiversity and ecosystem functions and services. In the run-up to the 13th Conference of the Parties on the 2020 Aichi biodiversity targets this year in Cancún, Mexico, we suggest how such assessments could be improved

and standardized globally.

The current state of ecosystems is measured by using direct observations. To detect trends in degradation or recovery over time, we must also define a baseline reference state. However, an arbitrarily chosen baseline, such as 50 years ago, might not reflect the true magnitude or direction of land degradation or recovery.

Instead, we suggest using an ecosystem's pre-degradation state, also known as its natural state. This state has no human-caused loss of biodiversity or of ecosystem functions. Contrasted with the ecosystem's current state, it provides a robust indicator of damage or recovery. It is determined by scientific research, but should not be viewed as a restoration target — those are merely political goals that balance social, economic and ecological interests.

A pre-degradation baseline could provide a universally applicable reference state. It reflects the deviation from genuine sustainability and, as a global standard, would ensure the fairness and comparability of ecosystem assessments across countries that are in different phases of economic development.

Janne S. Kotiaho *University of Jyväskylä, Finland.*
Ben ten Brink *PBL–Netherlands Environmental Assessment Agency, The Netherlands.*
Jim Harris *Cranfield University, Bedfordshire, UK.*
janne.kotiaho@jyu.fi

Holistic hydropower scheme for China

We propose three ways in which China could cut the human and environmental cost of its hydropower developments.

First, reform the laws that govern new dam developments in China. These should include provisions to address the impact of individual projects on river basins. Measuring and

understanding the cascade effects of dams and predicting ecosystem trends can guide decisions on fish propagation and optimal reservoir operation, for instance. Charges for hydroelectricity need to be rationalized and migrants' compensation standards updated. And inter-regional transfer payments to the authorities, intended to balance the social and environmental losses of affected areas with the gains in other areas, should correlate with the economic benefits of hydropower development.

Second, pay more attention to resettlement. Migrant populations could be relocated to mountainous areas along cascade reservoirs, which would boost infrastructure and preserve natural resources. People would need government help with land planning, education, employment and forming new social networks, without any devaluation of their indigenous knowledge, social capital or traditions.

Third, improve the governance of international rivers. China and its neighbours should share river management and water resources, cooperatively seeking solutions to the problems of hydropower production, environmental sustainability and climate change. China has a wealth of experience and technologies to contribute from its many previous hydropower projects.
Wenzhe Tang* *Tsinghua University, Beijing, China.*
twz@mail.tsinghua.edu.cn
**On behalf of 7 correspondents (see go.nature.com/h6xza4 for full list).*

CORRECTION

An article in the Nature Index 2016 Japan (*Nature* **531**, S127–S130; 2016) wrongly described an alumina film as having high thermal conductivity; in fact, it has high thermal insulation. It also misnamed JEOL Ltd as JOEL.

STELLAR ASTROPHYSICS

Supernovae in the neighbourhood

Detailed measurements of radioisotopes in deep-sea deposits, plus modelling of how they reached Earth, indicate that many supernovae have occurred near enough to have potentially influenced evolution. [SEE LETTERS P.69 & P.73](#)

ADRIAN L. MELOTT

For more than half a century, astronomers have speculated that supernovae have occurred close enough to Earth to affect the planet, possibly contributing to mass extinctions or climate change. In this issue, two studies of isotopes produced by supernovae reveal events thought to have happened a few million years ago: Wallner *et al.*¹ (page 69) report measurements of marine deposits that contain isotopes produced outside the Solar System, whereas Breitschwerdt *et al.*² (page 73) model the transport of iron-60 and use this to calculate the explosion times and sites of the likely sources. It is now possible to ask key questions with some precision; for example, could these supernovae have had substantial effects on Earth's climate and organisms — and perhaps even a role in human evolution?

The past 500 million years or so have witnessed mass-extinction events of varying intensity. These include the major end-Cretaceous event that wiped out the dinosaurs and about half of the planet's other species 66 million years ago, and a moderate event at the end of the Jurassic period 145 million years ago. In 1954, it was suggested³ that a nearby supernova caused the greatest mass-extinction event, which occurred about 250 million years ago (end-Permian) and affected more than 90% of species on Earth. For decades afterwards, progress in research centred on estimates of how often such stellar events might happen and what effects, if any, there might be on Earth and the interstellar medium. Unsurprisingly, these studies lacked precision.

Breakthrough computations⁴ in the 1990s showed that, for moderately close explosions — those occurring within about 100 parsecs (a few hundred light years) of Earth, expected every million years or so — radioactive nuclides can be transported and deposited on our planet. The isotopes are carried by dust grains that are only weakly affected by magnetic fields or gas pressure. Several

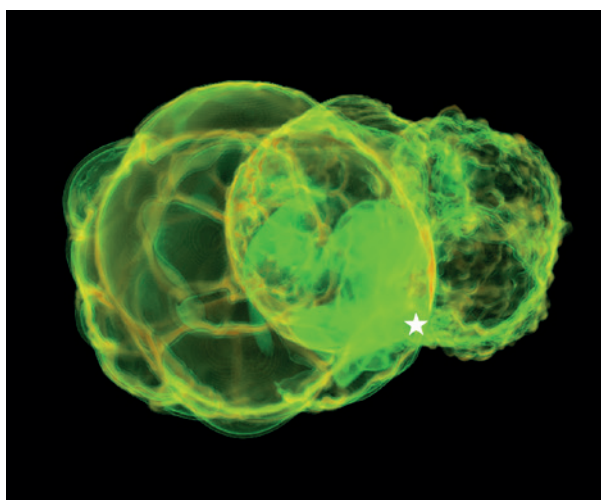


Figure 1 | Simulations of natural fireworks on a grand scale. Breitschwerdt *et al.*² performed high-resolution numerical simulations of supernovae that formed the Local Bubble (foreground, right) and the Loop I superbubble (left). Located in the Milky Way, these bubbles are large regions of hot, low-density gas surrounded by shells of swept-up debris that the Solar System has traversed for the past 5 million to 10 million years. In the image, the central solid green patch near Earth (represented by a white star) shows the mass distribution of the iron-60 isotope associated with the two bubbles 2.3 million years ago. The authors' simulations of the trajectory of iron-60 expelled from the supernovae agree with measurements of the isotope in the deep-sea crust by Wallner and colleagues¹.

candidate isotopes are problematic when used as evidence of supernovae: for example, beryllium-10 is also produced by radiation impinging on Earth's atmosphere, and plutonium-244 measurements suffer from background contamination originating from the testing of nuclear weapons. Nevertheless, concrete comparisons of estimates with measurements became possible for the first time.

In 1999, the detection⁵ of iron-60 in the deep-ocean ferromanganese crust broke the dam and led to a flood of results. Iron-60 is an excellent supernova indicator because it is produced abundantly in many types of supernova; non-supernova channels would produce only up to one-tenth as much. Nonetheless, the detection would not have been possible without the atom-counting capability of accelerator mass spectrometry, which allowed researchers to separate different isotopes by mass. This was followed by the first detailed computation⁶ of

the effects on Earth's atmosphere of photons and cosmic rays from supernova sources. That study⁶ established a 'kill distance' of roughly 8 parsecs, within which the effects of these high-energy particles on terrestrial biota would be catastrophic. Supernovae this close should, on average, occur as often as once every 800 million years or so⁷, occasionally interspersed with additional events due to γ -ray bursts.

Breitschwerdt *et al.* looked at the deposition history of iron-60 and the trajectories of stars from a likely precursor group — the Local Bubble, a region of hot gas that includes both the Solar System and the probable sites of the supernovae that formed the bubble (Fig. 1). They find that the nearest explosions were approximately 100 parsecs away. The authors highlight the main sources of uncertainty in their conclusions as being in the supernova yield of iron-60 and in its transport through the interstellar medium and deposition on Earth. Despite these uncertainties, their findings are an important proof of concept for a unified

picture of the dynamics of the interstellar medium, of the dynamics of the stellar group that gave rise to the supernova, and of the subsequent deposition of isotopes on Earth. The general set of parameters described in the study falls within a range suggested in recent work⁷ on analytical estimates of deposition processes.

By using accelerator mass spectrometry on three different deep-sea archives, encompassing four sediment cores, two iron-manganese crusts and two iron-manganese nodules, Wallner *et al.* greatly expand the amount of available data for iron-60, as well as for aluminium-26 and beryllium-10. They find iron-60 levels averaging at about 40 times background levels. These results favour supernovae as the source, in two main events: one at 1.7 million to 3.2 million years ago, and the other 6.5 million to 8.7 million years ago. Both events were longer in duration than could be explained by the passage of a single blast wave.

ZENTRUM ASTRON. ASTROPHYS., TU BERLIN, MICHAEL SCHULREICH

MICROBIOLOGY

Wallner and co-workers therefore propose that either a series of supernova events directly affected the Solar System, or that the Solar System passed through an interstellar medium polluted by the mixed products of multiple supernovae.

Either the direct or indirect picture is reasonable, given that the rate of supernovae is expected to vary significantly in the different environments traversed by the Solar System during its orbit about the Galaxy. The data provided by Wallner *et al.* will allow considerable refinement of models such as those of Breitschwerdt and colleagues. It will also facilitate research into several puzzling characteristics of the cosmic-ray spectrum (cosmic rays are mostly accelerated protons) that can be explained⁸ if a supernova occurred roughly 2 million years ago at the same distance from Earth (roughly 100 parsecs) as the supernovae detected by Breitschwerdt and colleagues.

What can we discern about Earth's violent past from this broadly unified picture? The recorded supernovae occurred beyond the kill distance, and no major global mass extinctions^{9,10} coincide with them. However, during the suggested interval there was a general decline in temperature that culminated in the extensive series of glaciations in the Pleistocene epoch (from about 2.6 million to 12,000 years ago), although we do not know if there is a link between supernova activity and colder temperature. This climatic variation may be one of the conditions that led to human evolution. Ionization of the atmosphere by supernovae may also lead to an increase in lightning¹¹ and possibly other climatic effects. The new studies^{1,2} will open up unexplored avenues of modelling and detailed investigation, providing deeper insight into what might have happened on Earth over the past 10 million years as a result of nearby stellar fireworks. ■

Adrian L. Melott is in the Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, USA.
e-mail: melott@ku.edu

1. Wallner, A. *et al.* *Nature* **532**, 69–72 (2016).
2. Breitschwerdt, D. *et al.* *Nature* **532**, 73–76 (2016).
3. Schindewolf, O. H. *Neues Jb. Geol. Paläontol.* **10**, 457–465 (1954).
4. Ellis, J., Fields, B. D. & Schramm, D. N. *Astrophys. J.* **470**, 1227–1236 (1996).
5. Knie, K. *et al.* *Phys. Rev. Lett.* **83**, 18–21 (1999).
6. Gehrels, N. *et al.* *Astrophys. J.* **585**, 1169–1176 (2003).
7. Fry, B. J., Fields, B. D. & Ellis, J. R. *Astrophys. J.* **800**, 71 (2015).
8. Kachelrieß, M., Neronov, A. & Semikoz, D. V. *Phys. Rev. Lett.* **115**, 181103 (2015).
9. Bambach, R. K. *Annu. Rev. Earth Planet. Sci.* **34**, 127–155 (2006).
10. Melott, A. L. & Bambach, R. K. *Paleobiology* **40**, 177–196 (2014).
11. Erlykin, A. D. & Wolfendale, A. W. *Surv. Geophys.* **31**, 383–398 (2010).

Fungus produces a toxic surprise

A protein fragment released by filaments of the fungus *Candida albicans* destroys host cells. This is the first demonstration that human fungal pathogens other than moulds can release toxic peptides. SEE ARTICLE P.64

AARON P. MITCHELL

The fungus *Candida albicans*, a common cause of infection in mucosal tissues, forms long filaments called hyphae, comprised of tubular cells, that are required for virulence in animals¹. Hyphal-associated adhesion proteins bind to host tissue, which is then degraded by hydrolytic enzymes. However, until now, no hyphal toxin that damages host cells has been identified. Because of this, the fungus is considered to be an ‘accidental’ pathogen that benignly inhabits our mucosal tissues, causing tissue damage by happenstance rather than design. That view must now change drastically. On page 64 of this issue, Moyes *et al.*² reveal a mechanism through which hyphae actively wage war on our cells — the production of a toxin that the authors call Candidalysin.

Interactions between *C. albicans* and epithelial cells, which line the body's cavities, occur during the early stages of mucosal infections such as thrush and vaginitis. Hyphae elicit³ several epithelial-cell responses, including the production of signalling molecules called cytokines that recruit cells of

the immune system to defend tissues, and loss of cell integrity through cell-membrane deterioration. Moyes *et al.* discovered that a strain of *C. albicans* in which the gene *ECE1* was mutated could not elicit epithelial-cell responses, despite growing apparently normal hyphae. Moreover, the authors validated these tissue-culture observations *in vivo* — the *ECE1* mutant was unable to reliably infect mucosa in a zebrafish swimbladder model and a mouse model of thrush.

ECE1 was one of the first genes to be identified in hyphal-specific expression screens more than 20 years ago⁴, yet until now it has been one of the most poorly understood genes in *C. albicans*. In fact, *ECE1* is among the most highly expressed genes in hyphae⁵, but its role has not previously been investigated thoroughly because mutants show no defects in hyphal morphology or cell proliferation^{4,6}. Thus, the function of the Ece1 protein has remained a puzzle.

How does Ece1 promote epithelial-cell responses? The protein's amino-acid sequence suggests that it is secreted from hyphae as a group of eight short protein fragments, or peptides, and so would be well positioned to

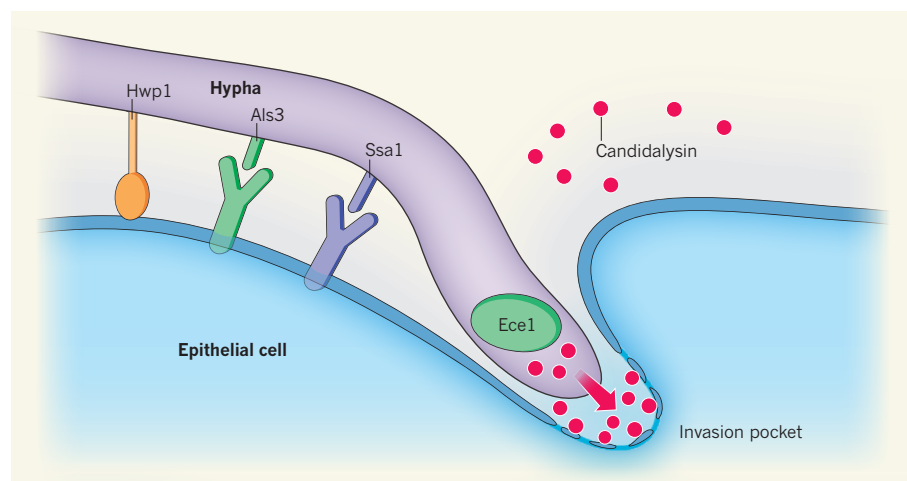


Figure 1 | A toxic relationship. The pathogenic fungus *Candida albicans* infects its host by forming filamentous structures called hyphae. Proteins on the hyphal surface — the adhesin Hwp1, the invasin Als3 and its partner Ssa1 — make contact with the host cell directly or through receptor proteins to promote adhesion and engulfment of the hypha by the host. Moyes *et al.*² report that the protein Ece1 is secreted from the hyphae as eight short peptides. One of these, Candidalysin, acts as a toxin that accumulates in the invasion pocket and attacks the host-cell membrane, leading to membrane permeabilization and the induction of host defences.

interact with host cells. Moyes and colleagues confirmed that all eight Ece1 peptides are secreted from hyphae. Analysis of synthetic versions of each peptide revealed that one, Ece1-III, elicits the same responses from epithelial cells as do hyphae. Moreover, precise deletion of the genetic region that codes for only Ece1-III created a mutant *C. albicans* that secreted the remaining seven peptides, but did not elicit epithelial-cell responses or cause mucosal disease in animal models. These results clearly demonstrate that Ece1-III mediates the pathogenic activity associated with *ECE1*.

By what mechanism does Ece1-III exert this activity? Certain chemical and structural features indicate that Ece1-III might function like peptide toxins, such as the bee-venom toxin melittin. Indeed, the authors show that the peptide causes rapid and transient permeabilization of artificial cell membranes *in vitro*. These activities are enhanced in the presence of cholesterol, a component of animal — but not fungal — membranes. The researchers conclude that Ece1-III acts as a peptide toxin, which they name Candidalysin.

Moyes and colleagues' study establishes that *C. albicans* hyphae evolved to damage host cells. When combined with our knowledge of hyphal adhesin proteins and enzymes, a simple program of tissue destruction emerges (Fig. 1). First, the hyphal-specific adhesin Hwp1 attaches to mucosal surfaces⁷. Second, the hyphal-specific invasion protein Als3, acting with the protein Ssa1, binds to receptors on the surface of the host cell, promoting engulfment of the hypha by the host cell⁸. Finally, Candidalysin accumulates in the invasion pocket around the hypha, attacking the host's cholesterol-containing membrane.

This attack leads to membrane permeabilization, leakage of cell contents and a defensive cytokine response, which serves to limit the size of the *C. albicans* population in healthy individuals. However, impaired defences in people with conditions such as AIDS, diabetes and some cancers permit *C. albicans* growth and consequent disease.

Every study raises fresh questions, and this one is no exception. For instance, why did a previous analysis⁶ find that *ECE1* was not necessary for invasive infection? The study in question was rigorous and efficiently tested many *C. albicans* genes using a mixture of mutant strains. One possible explanation, given that Ece1 functions extracellularly, is that an *ECE1* mutant was rescued by neighbouring cells that did express the gene.

Might other human pathogens of the genus *Candida* also produce peptide toxins? The *ECE1* gene is found in only a few other genomes. However, other aspects of *ECE1* might serve as good guides in the hunt for prospective toxin genes. For example, a small, highly expressed gene product with predicted chemical and structural features reminiscent of Ece1 would be an excellent candidate for a peptide toxin.

It remains unclear whether the seven Ece1 peptides that are secreted along with Candidalysin have a role in host interaction, or are just hitchhikers. Given that *C. albicans* spends most of its time as a commensal rather than a pathogenic fungus, it seems likely that Candidalysin or the other Ece1 peptides have roles in maintaining normal fungus–host interactions. Finally, Ece1 seems to be a promising drug target — inhibitors of its production or activity might be effective antifungal agents. Such an advance would be welcome, because there is currently a dearth of antifungal drugs, and the rates of disease and mortality associated with *C. albicans* infections are high⁹. ■

Aaron P. Mitchell is in the Department of Biological Sciences, Carnegie Mellon

University, Pittsburgh, Pennsylvania 15213, USA.

e-mail: apm1@cmu.edu

1. Mayer, F. L., Wilson, D. & Hube, B. *Virulence* **4**, 119–128 (2013).
2. Moyes, D. L. *et al. Nature* **532**, 64–68 (2016).
3. Naglik, J. R., Richardson, J. P. & Moyes, D. L. *PLoS Pathog.* **10**, e1004257 (2014).
4. Birse, C. E., Irwin, M. Y., Fonzi, W. A. & Sypherd, P. S. *Infect. Immun.* **61**, 3648–3655 (1993).
5. Bruno, V. M. *et al. Genome Res.* **20**, 1451–1458 (2010).
6. Noble, S. M., French, S., Kohn, L. A., Chen, V. & Johnson, A. D. *Nature Genet.* **42**, 590–598 (2010).
7. Staab, J. F., Bradway, S. D., Fidel, P. L. & Sundstrom, P. *Science* **283**, 1535–1538 (1999).
8. Filler, S. G. *Trends Microbiol.* **21**, 389–396 (2013).
9. Brown, G. D. *et al. Sci. Transl. Med.* **4**, 165rv113 (2012).

This article was published online on 30 March 2016.

ISLAND BIOGEOGRAPHY

Shaped by sea-level shifts

An analysis of changes in island topography and climate that have occurred since the last glacial maximum 21,000 years ago shows how sea-level change has influenced the current biodiversity of oceanic islands. [SEE LETTER P.99](#)

JOSÉ MARÍA FERNÁNDEZ-PALACIOS

Since the time of nineteenth-century naturalists such as Charles Darwin and Alfred Russel Wallace, islands have played a crucial part in the development of several scientific disciplines. Among these is the field of biogeography, which analyses the distributions of species and the historical, ecological and evolutionary processes behind them. In this issue, Weigelt *et al.*¹ (page 99) explore how the approximately 122-metre rise in sea level that occurred between the last glacial maximum (LGM), around 21,000 years ago, and the present day has contributed to the diversity of plant species seen on islands today.

The equilibrium theory of island biogeography (ETIB) was proposed in 1967 (ref. 2). This postulated that the species richness of an island depends both on the island's isolation, which controls the immigration rate of individual organisms from continents, and on its area, which controls the extinction rate of island species. This theory transformed biogeography from a mainly descriptive discipline to a quantitative and predictive one, and remains the most widely cited theory in the field.

The predictive strength of the ETIB was improved by embellishments such as the rescue effect³ and the target-area effect⁴, and we now know that the theory applies well to land-bridge islands — those that are separated from

the mainland by the current high sea level but are actually part of the continental shelf, such as Sumatra or Tasmania⁵. However, the predictions of the ETIB fail for volcanic islands — those that emerged as a result of the accumulation of magma on the oceanic crust and were never part of a continent, such as Hawaii or the Canary Islands. It also fails to work for micro-continents — islands that became detached from continents after the formation of a new oceanic ridge, such as Madagascar or New Zealand.

This failure is attributable mainly to the fact that the formation of new species (speciation) overrides immigration on volcanic islands and micro-continents⁶. Although these types of island may comprise less than 5% of existing islands, they contribute disproportionately to global biodiversity through their high numbers of endemic (exclusive) species.

In recent decades, there has been a shift from the idea of equilibrium in island biogeography to one of disequilibrium⁷. This thinking has been fuelled by the availability of increasingly complex and accurate bathymetric (underwater topography) charts, which have revealed extraordinary numbers of seamounts (underwater mountains) and guyots (table-shaped seamounts that were once islands but were later eroded and/or subsided beneath sea level). The general dynamic model⁸ proposes that the different developmental stages of a

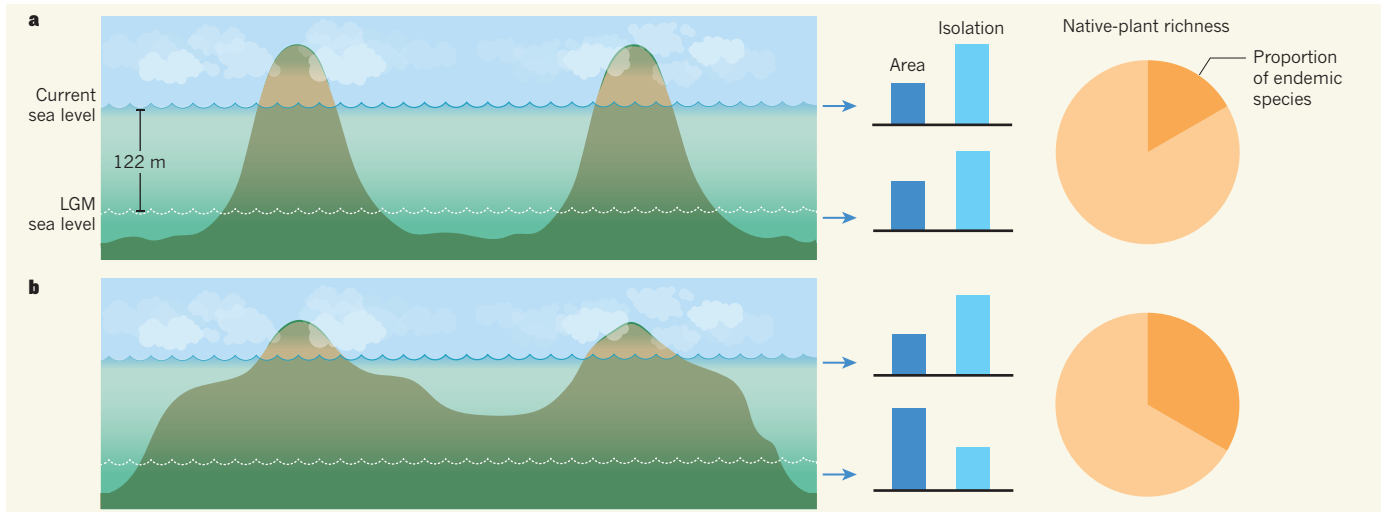


Figure 1 | Consequences of island dynamism. These hypothetical topographic configurations show how islands might have been affected by the 122-metre rise in sea level that has occurred since the last glacial maximum (LGM) 21,000 years ago. **a**, The two islands currently above sea level have not experienced substantial transformation since the LGM, such that their area and isolation have not shifted significantly as a result of the sea-level rise.

b, The two current islands formed a single and much larger entity during the LGM owing to the existence of an island platform. Sea-level rise caused that single island to undergo a fission event, resulting in a loss of area and increased isolation. Weigelt *et al.*¹ find that although the species richness of native plants is similar in both scenarios, the percentage of endemic species is larger in **b**, reflecting the islands' more complex and dynamic evolutionary history.

volcanic island control the biodiversity that it harbours.

Bathymetric charts have also enriched our understanding of how glacial cycles during the Pleistocene epoch (from around 2.6 million to 11,500 years ago) altered global sea-level height, and of the effect of these sea-level changes on island dynamism. These cycles have influenced islands' climates, and thus the altitudinal distribution of species and ecosystems. They have also contributed to changes in marine currents and wind regimes, which have affected connectivity, species sources and colonization rhythms⁹. Furthermore, glacial cycles affected the geographic configurations of islands, causing shifts in area, isolation and altitude, provoking island fissions and fusions, and inducing the emergence and submergence of seamounts¹⁰.

However, although there have been some attempts to clarify the role that sea-level shifts have had in shaping the biota of some volcanic archipelagos^{11–14}, the real consequences of sea-level oscillations for current island biodiversity remain obscure. Weigelt and colleagues have taken a key step towards filling this knowledge gap.

The authors analysed 184 oceanic islands worldwide, including volcanic islands and micro-continents, to see how they have changed since the LGM. The researchers assessed island area, isolation and elevation, the number of islands that existed, and the temperature and precipitation on the islands during the LGM and today. They also measured key components of the diversity of flowering plants (angiosperms), including the number of endemic and native species, and the proportion of endemics.

Weigelt *et al.* found that the number and

proportion of endemic angiosperm species found today is significantly higher on islands that were larger during the LGM than they are now (Fig. 1). However, they identified only weak effects of changes in temperature and precipitation on angiosperm diversity. The authors argue that islands that were repeatedly larger in the past (low-sea-level events occurred several times in the Pleistocene¹⁵) might have been sites of within-island speciation that has left a strong imprint on endemism today. By contrast, species richness was mostly related to islands' current environmental characteristics, suggesting an equilibrium with the present conditions.

Contrary to the significant effect of change in island area on endemic species, Weigelt *et al.* found that shifts in island isolation due to sea-level rise had no effect on the richness of native or endemic species, nor on the proportion of endemics. They observe that the difference between current island isolation and that at the LGM is small, because most oceanic islands are quite isolated and, consequently, sea-level shifts have not significantly varied their degree of isolation.

The authors conclude that considering environmental change during the Late Quaternary period (the past 130,000 years) is essential to understanding current patterns of island endemism and its underlying evolutionary dynamics. Furthermore, as the authors rightly point out, future research is needed on other, still poorly understood glaciation-associated shifts, for instance, in ocean currents, wind regimes and species source areas; on the emergence and disappearance of island-summit ecosystems and their ecological and evolutionary consequences; and on the part

that the reiterative emersion and submergence of seamounts has played in the past colonization of nearby islands. Such knowledge will improve our understanding of how current global warming will affect one of the planet's vital reservoirs of biodiversity — the oceanic islands. ■

José María Fernández-Palacios is in the *Island Ecology and Biogeography Group, Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Universidad de La Laguna (ULL), La Laguna 38200, Tenerife, Canary Islands, Spain. e-mail: jmferpal@ull.es*

1. Weigelt, P., Steinbauer, M. J., Sarmiento Cabral, J. & Kreft, H. *Nature* **532**, 99–102 (2016).
2. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
3. Brown, J. H. & Kodric-Brown, A. *Ecology* **58**, 445–449 (1977).
4. Lomolino, M. V. *Oikos* **57**, 297–300 (1990).
5. Whittaker, R. J. & Fernández-Palacios, J. M. *Island Biogeography: Ecology, Evolution and Conservation* 2nd edn (Oxford Univ. Press, 2007).
6. Lomolino, M. V. *Glob. Ecol. Biogeogr.* **9**, 1–6 (2000).
7. Heaney, L. R. *Glob. Ecol. Biogeogr.* **9**, 59–74 (2000).
8. Whittaker, R. J., Triantis, K. A. & Ladle, R. J. *J. Biogeogr.* **35**, 977–994 (2008).
9. Bellemain, E. & Ricklefs, R. E. *Trends Ecol. Evol.* **23**, 461–468 (2008).
10. Fernández-Palacios, J. M. *et al. Glob. Ecol. Biogeogr.* <http://dx.doi.org/10.1111/geb.12320> (2015).
11. Price, J. P. & Elliott-Fisk, D. L. *Pacif. Sci.* **58**, 27–45 (2004).
12. Ali, J. R. & Aitchison, J. C. *J. Biogeogr.* **41**, 1227–1241 (2014).
13. Warren, B. H., Strasberg, D., Bruggemann, J. H., Prys-Jones, R. P. & Thébaud, C. *Cladistics* **26**, 526–538 (2010).
14. Rijdsdijk, K. F. *et al. J. Biogeogr.* **41**, 1242–1254 (2014).
15. Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. *Nature* **437**, 125–128 (2005).

This article was published online on 30 March 2016.

CLIMATE SCIENCE

Water's past revisited to predict its future

A reconstruction of 1,200 years of water's history in the Northern Hemisphere, based on proxy data, fuels the debate about whether anthropogenic climate change affected twentieth-century precipitation. [SEE LETTER P.94](#)

MATTHEW E. KIRBY

Predicting future climate is one of science's great challenges¹. Climate models are key to these predictions, but such models must be validated through comparison with, and integration of, measured or proxy-based climate data². On page 94 of this issue, Ljungqvist *et al.*³ take on this challenge by reconstructing the past 12 centuries of water history in the Northern Hemisphere from proxy records. Their reconstruction for the twentieth century diverges sharply from climate-modelling results, and calls for greater development of water-sensitive proxy data and their integration into modelling efforts.

Estimating future water availability in a warming world is crucial, for various socio-economic and ecological reasons⁴. To compensate for the absence of measured data for future times, climate models are often assessed using twentieth-century data and, in many cases, inferred data from periods that pre-date climate measurements. Palaeoclimate data are derived from proxies of measured data, through the analysis of climate-sensitive archives such as ice cores, lake sediments and mineral deposits in caves (Fig. 1).

Ljungqvist and colleagues undertook a formidable task in this area of research. They compiled various water- and temperature-sensitive proxy data for the past 1,200 years, assessed age controls for the data, compared each data set across regions and continents, and then compared their proxy results with previously reported climate-model results. Their reconstruction of hydroclimate is generally consistent with model results for the period concerned — for example, it indicates a higher percentage

of wetter-than-average land (relative to the average before the twentieth century) during the ninth to eleventh and the twentieth centuries, and the opposite during the twelfth to nineteenth centuries. Persistent precipitation dipoles (north–south patterns of precipitation anomalies that have opposing signs), such as that observed now in the western United States⁵, are also evident. However, the data do not capture the climate trends⁶ known as

the Little Ice Age and the Medieval Climatic Anomaly.

Perhaps the most notable finding is that the proxy-based reconstruction diverges from model outputs for the twentieth century. Specifically, the proxy results indicate no difference in twentieth-century water dynamics compared with the pre-industrial era, whereas model results indicate the development of a substantial bimodal distribution in precipitation between extreme wetness and extreme dryness — dry gets drier and wet gets wetter (DDWW).

The DDWW idea is not new⁷. This concept is based largely on the expectation that higher global temperatures will intensify the water cycle, increasing the average rainfall and/or rainfall intensity, as well as amplifying net evaporation. This intensification is related to the Clausius–Clapeyron equation⁷, which states that moisture content in the lowest part of the atmosphere will rise with higher temperatures. Already-wet regions might therefore receive more-frequent and/or more-intense precipitation. Conversely, dry regions will become drier as warmer temperatures expand the subtropical dry zones and intensify areas of high pressure (low precipitation). But this picture is an oversimplification and effects are likely to manifest in unforeseen ways⁸.

Whether twentieth-century water-cycle intensification and the associated DDWW response have been detected is still controversial⁹. The divergence of Ljungqvist and colleagues' hydroclimate reconstruction from model results for the twentieth century certainly adds fuel to the fiery debate. Do their results invalidate current predictive models? Certainly not. But they do highlight a big challenge for climate modellers, and present major research opportunities both for modellers and for climate scientists who work with proxy data¹⁰.

For example, the authors highlight the continued importance of combining proxy-based results with climate models. This is more easily said than done. Not every site contains the same types of proxy, and the response of sources of proxy data to climate can change over time. Moreover, the climatic sensitivity of proxy-data sources varies from type to type and site to site.

To address these limitations, palaeoclimate scientists must



Figure 1 | Calcite columns in Lechuguilla Cave, New Mexico. Cave mineral deposits can provide proxy data on past precipitation patterns. Ljungqvist *et al.*³ have reconstructed the water history of the Northern Hemisphere for the past 1,200 years, based on proxies, including mineral deposits.

MICHAEL NICHOLS/NATIONAL GEOGRAPHIC CREATIVE

standardize best practices for proxy sampling, reconstructions, dating and statistical analyses. One priority should be to select key sites for proxy sampling and analysis. For example, are there crucial geographical gaps that require attention and that strike an urgent socio-economic or ecological chord^{4,11}? Fortunately, several initiatives are already focusing on improving efforts to integrate models and proxy results, for example the Paleoclimate Modelling Intercomparison Project and the Past Global Changes project.

Ljungqvist *et al.* were, of course, constrained by the data available for analysis — indeed, their efforts reveal a shocking lack of data. For example, Figure 1 of their paper³ highlights the vast geographical gaps between proxy sites. Immense areas of the Northern Hemisphere still require exploration for proxy development, many in highly populated regions. The current analysis should therefore be revisited

as proxy records from these regions become available.

Nevertheless, this research is a crucial first step in the use of models and proxy data to reconstruct and explain the history of water — and not just of temperature change — through time. Future research efforts should improve, test and extend Ljungqvist and colleagues' results. Global warming will undoubtedly change Earth's water cycle, so the more that is known about the cycle's past behaviour through proxy records, and the better those changes can be modelled, the more confidence we will have in predictive models as we forge ahead into the twenty-first century. ■

Matthew E. Kirby is in the Department of Geological Sciences, California State University, Fullerton, Fullerton, California 92834, USA.
e-mail: mkirby@fullerton.edu

1. Kirtman, B. *et al.* in *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 953–1028 (Cambridge Univ. Press, 2013).
2. Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T. & Meko, D. M. *J. Clim.* **27**, 7529–7549 (2014).
3. Ljungqvist, F. C. *et al.* *Nature* **532**, 94–98 (2016).
4. Conway, D. *et al.* *Nature Clim. Change* **5**, 837–846 (2015).
5. Wise, E. K. *Geophys. Res. Lett.* **37**, L07706 (2010).
6. Mann, M. E. *et al.* *Science* **326**, 1256–1260 (2009).
7. Trenberth, K. E., Dai, A., Rasmussen, R. M. & Parsons, D. B. *Bull. Am. Meteorol. Soc.* **84**, 1205–1217 (2003).
8. Cook, B. I., Smerdon, J. E., Seager, R. & Coats, S. *Clim. Dyn.* **43**, 2607–2627 (2014).
9. Prein, A. F., Holland, G. J., Rasmussen, R. M., Clark, M. P. & Tye, M. R. *Geophys. Res. Lett.* **43**, 1272–1279 (2016).
10. Schmidt, G. A. J. *Quat. Sci.* **25**, 79–87 (2010).
11. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. *Nature* **403**, 853–858 (2000).

NEUROSCIENCE

Untangling autism

A clever dissection of the roles of the *Ptchd1* gene in the brains of mice demonstrates one way to untangle the complex relationships between the causes and symptoms of neurodevelopmental disorders. SEE ARTICLE P.58

SCOTT BOLKAN & JOSHUA A. GORDON

Our current understanding of neurodevelopmental disorders can be thought of as a tangled mess of threads. The clinical presentations of such disorders are so variable that people can receive the same diagnosis despite not sharing a single symptom. Furthermore, the underlying risk factors — be they genetic or environmental — can be associated with not just one, but several disorders. Untangling these threads to link cause and outcome seems a Sisyphean task, but it is a crucial one if we are to develop treatment approaches that have a solid neurobiological basis. In this issue, Wells *et al.*¹ (page 58) use a mouse model of a human genetic condition to grasp the loose end of a single thread and carefully extricate it from the tangled mess. In doing so, they map a precise set of genetically linked symptoms onto dysfunction of a neuronal structure that gates the flow of information across multiple brain circuits.

Wells and colleagues' approach is based on clinical observations^{2,3} of a link between autism spectrum disorder and mutations in the *PTCHD1* gene. Autism spectrum disorder involves a tremendously debilitating disruption of social and cognitive function, and is frequently associated with several neurodevelopmental diagnoses that have overlapping symptoms, including intellectual

disability and attention deficit hyperactivity disorder.

As is typical for genes associated with neurodevelopmental conditions, *PTCHD1* mutations are seen in only a small fraction of people with autism spectrum disorder⁴. Not everyone with the risk-associated gene variant develops the disorder, but the mutation does raise disease risk substantially: more than 40% of individuals with the mutation develop autism-like behaviours, compared with about 1% of the general population⁵. Nonetheless,

how a single gene can alter brain function to produce particular symptoms remains unclear.

Wells *et al.* began to unravel these issues by generating mice that produce a truncated, non-functional form of the *PTCHD1* protein. The authors report that *Ptchd1* mutant mice exhibit behavioural abnormalities that are largely consistent with those seen in people with autism spectrum disorder, including sleep disruption, hyper-aggression and deficits in attention and learning.

In seeking to understand how the *Ptchd1* mutation can produce such a broad set of symptoms, the authors were struck by the observation that, during early postnatal development, expression of the gene is highly enriched in a neuronal structure called the thalamic reticular nucleus (TRN) in the thalamus region of the brain. This structure sends inhibitory neuronal projections to all areas of the thalamus, and thus exerts powerful control over the flow of neuronal activity within and across brain circuits that process information

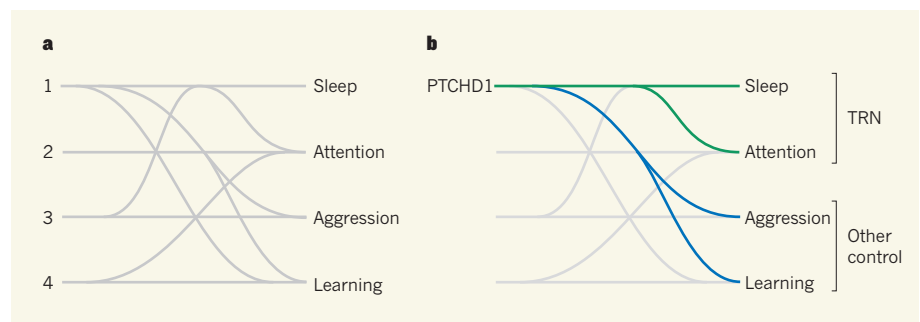


Figure 1 | Tracing the threads between causes and symptoms of autism. **a**, Autism spectrum disorder is caused by a range of environmental and genetic factors (symbolized here by numbers), each of which can lead to one or more of many symptoms. Untangling which causes lead to which symptoms is a major challenge. **b**, Wells *et al.*¹ have done just that for the *PTCHD1* protein, which is mutated in some people with autism spectrum disorder. They find that, in mice, *PTCHD1* regulates sleep, attention, aggression and learning. Furthermore, sleep and attention, but not aggression and learning, are mediated by *PTCHD1* expression in a neuronal structure called the thalamic reticular nucleus (TRN).

about vision, movement, cognition and more. The TRN is known as the 'guardian' or 'gate-keeper' of activity in these thalamic brain circuits and has regulatory roles in sleep⁶ and attentional processes^{7,8}. As such, dysfunction in the TRN could plausibly produce wide-ranging behavioural abnormalities.

Focusing on the TRN, Wells *et al.* traced a route from gene to behaviour in *Ptchd1* mutant mice. The authors monitored the activity of TRN neurons, and found fewer bursts of activity in mutant animals than in controls. These changes were attributable to a drastic reduction in the activity of the SK-channel protein, which mediates passage of potassium ions across the cell membrane and normally promotes activity bursts. With reduced levels of channel activity, and reduced numbers of bursts, TRN neurons fail to properly inhibit activity in other regions of the thalamus, including the lateral geniculate nucleus — the main visual relay between the eyes and the brain's visual cortex. This failure to properly inhibit thalamic relays seems to disrupt visual attention, among other behaviours.

Which of the wide-ranging abnormalities seen in the *Ptchd1* mutant mice truly depend on this disruption of TRN function? To answer this question, Wells *et al.* generated mice that lacked the *Ptchd1* gene exclusively in the TRN. In a remarkable dissociation of symptoms, they found that TRN-restricted *Ptchd1* mutant mice display hyperactivity and deficits in sleep and attention, but do not exhibit the learning deficits and hyper-aggression observed in mice carrying the brain-wide mutation (Fig. 1). Moreover, using a pharmacological agent to restore SK-channel function in adult mice, Wells *et al.* were able to re-establish activity levels in the thalamic relays controlled by the TRN. This reversed abnormal hyperactivity and deficits in sleep and attention, but not impaired learning or hyper-aggression. Successful restoration of normal sleep and attention in adult animals raises hopes that treatments targeting the SK channel might benefit people with *PTCHD1* mutations, and perhaps also those with autism spectrum disorder caused by other factors.

The authors' study indicates that both TRN-specific deletion of *PTCHD1* and systemic enhancements in SK-channel function affect the same subset of behavioural outcomes. This consistency strongly supports the idea that *PTCHD1* exerts its effects on TRN function (and therefore on activity, sleep and attention) through the SK channel. However, the findings also argue that the effects of *PTCHD1* on learning and aggression are caused by an altogether different mechanism that both acts outside the TRN and is independent of SK-channel function. Determining which brain regions and biochemical pathways might be responsible for these behavioural changes will probably be important for fully reversing the deficits

caused by *PTCHD1* mutations.

Wells *et al.* have placed their fingertips on the loose thread of *PTCHD1* expression, following it to the TRN, through the SK channel and to its end in behaviours associated with neurodevelopmental disease. In doing so, they have identified the SK channel as a potential pharmacological target for treating autism spectrum disorder, while simultaneously highlighting the usefulness of animal models of human genetic conditions for studying neurodevelopmental and psychiatric diseases. Although the role of *PTCHD1* in learning and aggression remains elusive, the current paper shows one way to navigate the daunting complexity of neuronal-circuit function to reveal the mechanisms by which gene dysfunction leads to changes in behaviour. ■

MOLECULAR BIOLOGY

Breaks in the brain

A high-throughput approach has found clusters of DNA double-strand breaks in neural cells. Most of the clusters are in large genes that are associated with neural function, which suggests that the breaks may have tissue-specific roles.

THOMAS W. GLOVER & THOMAS E. WILSON

Studies over the past decade have revealed a surprising degree of structural variation in human genomes. Structural variants (SVs) in germline DNA are now known to be a major factor in normal genomic variation and an important class of mutation in genomic disorders, and they arise frequently in cancers. Many SVs are thought to result from DNA-replication errors¹, so they would be expected to occur at a high frequency in dividing somatic cells (those that do not undergo meiotic division). However, in contrast to heritable SVs in germline cells, little is known about somatic SVs and their impact on tissue function and disease. Writing in *Cell*, Wei *et al.*² provide insight into these questions by examining the landscape of DNA double-strand breaks (DSBs) that arise in mouse neural cells.

The difficulty in studying somatic SVs arises primarily from technical challenges in detecting rare events in cell populations. Sensitive sequencing technologies coupled with bioinformatic tools have begun to provide glimpses of somatic SVs *in vivo*, especially in the brain, where up to 40% of individual neurons have been found to contain megabase-scale copy-number variations (CNVs; a form of SV in which the number of copies of a genomic region varies between cells or individuals)^{3,4}. However, because of the low resolution of many approaches, the true prevalence of SVs in neurons or other cells, and the mechanisms

Scott Bolkan is in the Department of Neuroscience, Columbia University, New York, New York 10032, USA. Joshua A. Gordon is in the Department of Psychiatry, Columbia University.
e-mail: jg343@cumc.columbia.edu

1. Wells, M. F., Wimmer, R. D., Schmitt, L. I., Feng, G. & Halassa, M. M. *Nature* **532**, 58–63 (2016).
2. Coe, B. P., Girirajan, S. & Eichler, E. E. *Curr. Opin. Neurobiol.* **22**, 829–836 (2012).
3. Pinto, D. *et al. Nature* **466**, 368–372 (2010).
4. Noor, A. *et al. Sci. Transl. Med.* **2**, 49ra68 (2010).
5. Chaudhry, A. *et al. Clin. Genet.* **88**, 224–233 (2015).
6. Barthó, P. *et al. Neuron* **82**, 1367–1379 (2014).
7. Wimmer, R. D. *et al. Nature* **526**, 705–709 (2015).
8. McAlonan, K., Cavanaugh, J. & Wurtz, R. H. *Nature* **456**, 391–394 (2008).

This article was published online on 23 March 2016.

by which they arise, are unknown.

Wei *et al.* used a sensitive, targeted assay known as high-throughput genomic translocation sequencing. This method allows genome-wide detection of naturally occurring 'prey' DSBs using experimentally induced 'bait' DSBs targeted elsewhere in the genome. The two DSBs are joined by cellular DNA-repair processes, leading to a translocation between the genomic regions. Sequencing of the resulting breakpoint junction allows mapping and characterization of the DSBs at nucleotide-level resolution.

The authors first created DSBs at bait loci on three mouse chromosomes in cultured neural stem/progenitor cells (NSPCs) that lacked the protein *Xrcc4*, which is essential for a DSB-repair process called non-homologous end-joining (NHEJ); preventing this joining enriches for cells with rearrangements. The cells also lacked the protein *p53*; this lack promotes cell survival. The researchers identified thousands of prey DSBs, with 61% located close to the bait DSBs and the rest spread across the genome (Fig. 1). Strikingly, many were found within three recurrent DSB clusters (RDCs); two of these were in the *Lsmp* and *Npsa3* genes, which are unusually large genes that encode a neural-cell-specific adhesion molecule and a transcription factor, respectively. The third RDC reflected DSBs that occurred close to the bait.

When the authors applied the same procedure in mouse B cells (a type of immune cell), they found B-cell-specific RDCs but no

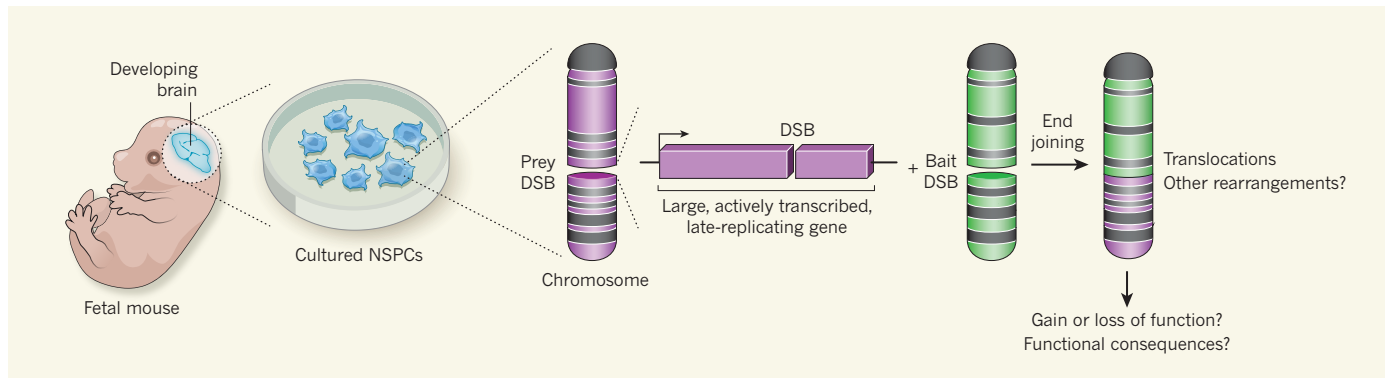


Figure 1 | Translocation hotspots. Wei *et al.*² performed an unbiased genome-wide screen for DNA double-strand breaks (DSBs) in cultured mouse neural stem/progenitor cells (NSPCs). Their assay identifies prey DSBs that have joined with experimentally induced bait DSBs to generate inter-chromosomal translocations. The researchers found hotspots for recurrent DSBs in large, actively transcribed, late-replicating genes that function in synapses between neural cells or in neural adhesion. Many more recurrent DSB clusters were identified when the cells were subjected to replication stress. These DSBs could lead to genomic rearrangements that contribute to genetic heterogeneity in different cells of the brain or other tissues, giving rise to functional consequences *in vivo*.

translocations involving *Lsmp*. Transcription analysis showed *Lsmp* expression in NSPCs but not B cells, suggesting that the susceptibility of certain genetic regions to DSBs in certain cell types is mechanistically linked to transcription.

The authors then examined the impact of DNA-replication stress on DSBs by treating NSPCs with low doses of aphidicolin, an inhibitor of DNA polymerase enzymes that is widely used to study common fragile site (CFS) instability⁵. This approach identified 24 additional RDCs, all in genes, and more than 300 candidate RDCs. In total, 26 of the 27 RDCs were in genes larger than 400 kilobases, and 13 were in genes larger than 1 megabase, including many CFS genes. More than 90% (24 out of 26) of the RDC genes in NSPCs are involved in neural adhesion or the function of synapses (connections between neurons), and many are implicated in neurodevelopmental and neuropsychiatric disorders, including *Lsmp*, *Nrx1* and *Dcc*. All but one of the RDC genes were actively transcribed, and all but one completed replication late in the S phase of the cell cycle. Comparing the sequences at breakpoint junctions between NHEJ-proficient and NHEJ-deficient cells indicated that both canonical and alternative end-joining pathways mediate the translocations.

These findings of large gene size, active transcription, late replication and induction by replication stress strongly suggest that interference between transcription and replication contributes to the high frequency of DSBs and genomic rearrangements in NSPCs. An earlier study⁵ linked these same factors to CNVs and CFSs in human fibroblast cells and mouse embryonic stem cells. The study found both spontaneous and replication-stress-induced CNVs of sizes ranging from 1 kilobase to several megabases across the genome, but with 'hotspots' in large, late-replicating, actively transcribed genes, including *LSAMP* in human fibroblasts and other genes with neural functions. These hotspots were also CFSs — regions

in which gene structure and slowed or stalled replication led to the passage of unreplicated DNA into mitotic cell division^{6,7}.

Strikingly, hotspot CNVs clustered in the centre of large genes in a pattern similar to the distribution of RDC breakpoints identified by Wei and colleagues. This instability was again evident only when the genes were transcribed. There was also considerable overlap in the genes identified in these two studies, with 5 out of 9 (56%) of the human CNV hotspots corresponding to RDC genes and 52% of the RDC genes showing CNVs in mouse embryonic stem cells. A key insight from Wei and colleagues' results is that instability at these sites is associated with DSBs. In addition to giving rise to translocations, the DSBs detected by Wei *et al.*, which probably arise from stalled replication forks that form during DNA synthesis, might also lead to CNVs through end-joining or other DNA-repair pathways. Understanding the disposition of transcription-dependent DNA lesions between different repair pathways and their coordination in the cell cycle are challenges for future studies.

In addition to mechanistic insights, Wei and colleagues' findings have implications for our understanding of genome stability in the brain and other tissues. It has been speculated that somatic genomic rearrangements contribute to neuronal diversity, learning and memory⁸, processes in which many large RDC-associated genes function. Studies of SVs in placenta, skin and other tissues have also suggested essential roles in cell biology^{9,10}. The research group behind the current study previously showed¹¹ that NHEJ is essential for NSPC differentiation, which implies a role for NHEJ-mediated DSB repair during normal brain development. It is possible that subpopulations of cells in the brain display the same types of rearrangements, and that some are positively selected. Similar studies *in vivo* and in other tissues are needed to test this hypothesis.

The fact that many of the largest mammalian genes are transcribed in the brain and have

neuronal functions supports the potential for brain-specific SVs. It is enticing to speculate that programmed SV formation could explain why large gene structure is evolutionarily conserved despite being potentially unstable. However, an alternative explanation is length-dependent fine-tuning of gene expression through protein binding to brain-specific DNA modifications¹². We also need to understand what other factors contribute to DSBs and SV formation at large transcribed genes and why some are more susceptible than others. Finally, knowledge of the full spectrum of SVs found in the brain and other tissues will be required to discern their potential consequences. Many *de novo* SVs are likely to be deleterious, but variants that lead to gained or altered functions would be of immense interest. ■

Thomas W. Glover and Thomas E. Wilson are in the Departments of Human Genetics and Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.
e-mails: glover@med.umich.edu; wilsonte@med.umich.edu

1. Lee, J. A., Carvalho, C. M. B. & Lupski, J. R. *Cell* **131**, 1235–1247 (2007).
2. Wei, P.-C. *et al.* *Cell* **164**, 644–655 (2016).
3. Cai, X. *et al.* *Cell Rep.* **8**, 1280–1289 (2014).
4. McConnell, M. J. *et al.* *Science* **342**, 632–637 (2013).
5. Wilson, T. E. *et al.* *Genome Res.* **25**, 189–200 (2015).
6. Durkin, S. G. & Glover, T. W. *Annu. Rev. Genet.* **41**, 169–192 (2007).
7. Minocherhomji, S. *et al.* *Nature* **528**, 286–290 (2015).
8. Muotri, A. R. & Gage, F. H. *Nature* **441**, 1087–1093 (2006).
9. Hannibal, R. L. *et al.* *PLoS Genet.* **10**, e1004290 (2014).
10. O'Huallachain, M., Karczewski, K. J., Weissman, S. M., Urban, A. E. & Snyder, M. P. *Proc. Natl Acad. Sci. USA* **109**, 18018–18023 (2012).
11. Gao, Y. *et al.* *Cell* **95**, 891–902 (1998).
12. Gabel, H. W. *et al.* *Nature* **522**, 89–93 (2015).

This article was published online on 23 March 2016.

Climate-smart soils

Keith Paustian^{1,2}, Johannes Lehmann³, Stephen Ogle^{2,4}, David Reay⁵, G. Philip Robertson⁶ & Pete Smith⁷

Soils are integral to the function of all terrestrial ecosystems and to food and fibre production. An overlooked aspect of soils is their potential to mitigate greenhouse gas emissions. Although proven practices exist, the implementation of soil-based greenhouse gas mitigation activities are at an early stage and accurately quantifying emissions and reductions remains a substantial challenge. Emerging research and information technology developments provide the potential for a broader inclusion of soils in greenhouse gas policies. Here we highlight 'state of the art' soil greenhouse gas research, summarize mitigation practices and potentials, identify gaps in data and understanding and suggest ways to close such gaps through new research, technology and collaboration.

Evidence points to agriculture as the first instance of human-caused increases in greenhouse gases (GHGs), several thousand years ago¹. Agriculture and associated land-use change remain a source for all three major biogenic GHGs: carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O). Land use contributes ~25% of total global anthropogenic GHG emissions: 10%–14% directly from agricultural production, mainly via GHG emissions from soils and livestock management, and another 12%–17% from land cover change, including deforestation^{2,3}. Although soils contribute a major share (37%; mainly as N₂O and CH₄) of agricultural emissions³, improved soil management can substantially reduce these emissions and sequester some of the CO₂ removed from the atmosphere by plants, as carbon (C) in soil organic matter (in this Perspective, our discussion of soil C refers solely to organic C). In addition to decreasing GHG emissions and sequestering C, wise soil management that increases organic matter and tightens the soil nitrogen (N) cycle can yield powerful synergies, such as enhanced fertility and productivity, increased soil biodiversity, reduced erosion, runoff and water pollution, and can help buffer crop and pasture systems against the impacts of climate change⁴.

The inclusion of soil-centric mitigation projects within GHG offset markets^{5,6} and new initiatives to market 'low-carbon' products⁷ indicate a growing role for agricultural GHG mitigation⁸. Moreover, interest in developing aggressive soil C sequestration strategies has been heightened by recent assessments, which project that substantial terrestrial C sinks will be needed to supplement large cuts in GHG emissions to achieve GHG stabilization levels of 450 parts per million CO₂ equivalent or below, consistent with the goal of a mean global temperature increase of less than 2 °C (ref. 9). Soil C sequestration is one of a few strategies that could be applied at large scales⁹ and potentially at low cost; as an example, the French government has proposed¹⁰ to increase soil C concentration in a large portion of agricultural soils globally, by 0.4% per year, in conjunction with the Conference of the Parties to the UN Framework Convention on Climate Change (UNFCCC) negotiations in December 2015. This would produce a C sink increase of 1.2 petagrams (Pg) of C per year (ref. 10).

An extensive body of field, laboratory and modelling research over many decades demonstrates that improved land use and management practices can reduce soil GHG emissions and increase soil C stocks. However, implementing effective soil-based GHG mitigation strategies on a large scale will require the capacity to measure and monitor GHG reductions with acceptable accuracy, quantifiable uncertainty and

at relatively low cost. Targeted research to improve predictive models, expanded observational networks to support model validation and uncertainty bounds, 'Big Data' approaches to integrate land use, management and environmental drivers, and technologies with which to engage actively with land users at the grass-roots level, are all key elements in realizing the potential GHG mitigation from climate-smart agricultural soils.

Process controls and mitigation practices

Soil C sequestration via improved management

Soils constitute the largest terrestrial organic C pool (~1,500 Pg C to a depth of 1 m; 2,400 Pg C to 2 m depth¹¹), which is three times the amount of CO₂ currently in the atmosphere (~830 Pg C) and 240 times the current annual fossil fuel emissions (~10 Pg)⁹. Thus, increasing net soil C storage by even a few per cent represents a substantial C sink potential.

Proximal controls on the soil C balance include the rate of C addition as plant residue, manure or other organic waste, minus the rate of C loss (via decomposition). Hence, C stocks can be increased by increasing organic matter inputs or by reducing decomposition rates (for example, by reducing soil disturbance), or both, leading to net removal of C from the atmosphere¹². However, soil C accrual rates decrease over time as stocks approach a new equilibrium. Therefore net CO₂ removals are of limited duration, often attenuating after two to three decades¹³.

Unmanaged forests and grasslands typically allocate a large fraction of their biomass production belowground and their soils are relatively undisturbed; accordingly, native ecosystems usually support much higher soil C stocks than their agricultural counterparts, and soil C loss (typically 0.5 to >2 megagrams (Mg) of C per hectare per year) following land conversion to cropland has been extensively documented^{14,15}. Total losses once the soil approaches a new equilibrium are typically ~30%–50% of topsoil (for example, 0–30 cm depth) C stocks¹⁵. Hence, avoiding conversion and degradation of native ecosystems is a strong mitigation alternative. Conversely, restoration of marginal or degraded lands to perennial forest or grassland increases soil C storage (Fig. 1), although usually at a slower rate than the original conversion losses^{16,17}. Restoring wetlands that have been drained for agricultural use reduces ongoing decomposition losses, which can be as high as 5–20 Mg C ha⁻¹ yr⁻¹ (ref. 18), and can also restore C sequestration (Fig. 1), though methane emissions may increase^{19,20}. Land-use conversions may, however, conflict with agricultural production and food security objectives, entailing the need for a broad-based accounting of net GHG implications (Table 1)²¹.

¹Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, USA. ²Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado, USA.

³Atkinson Center for a Sustainable Future, Department of Soil and Crop Sciences, Cornell University, Ithaca, New York, USA. ⁴Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, Colorado, USA. ⁵School of Geosciences, University of Edinburgh, Edinburgh, UK. ⁶W. K. Kellogg Biological Station and Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan, USA. ⁷Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, UK.

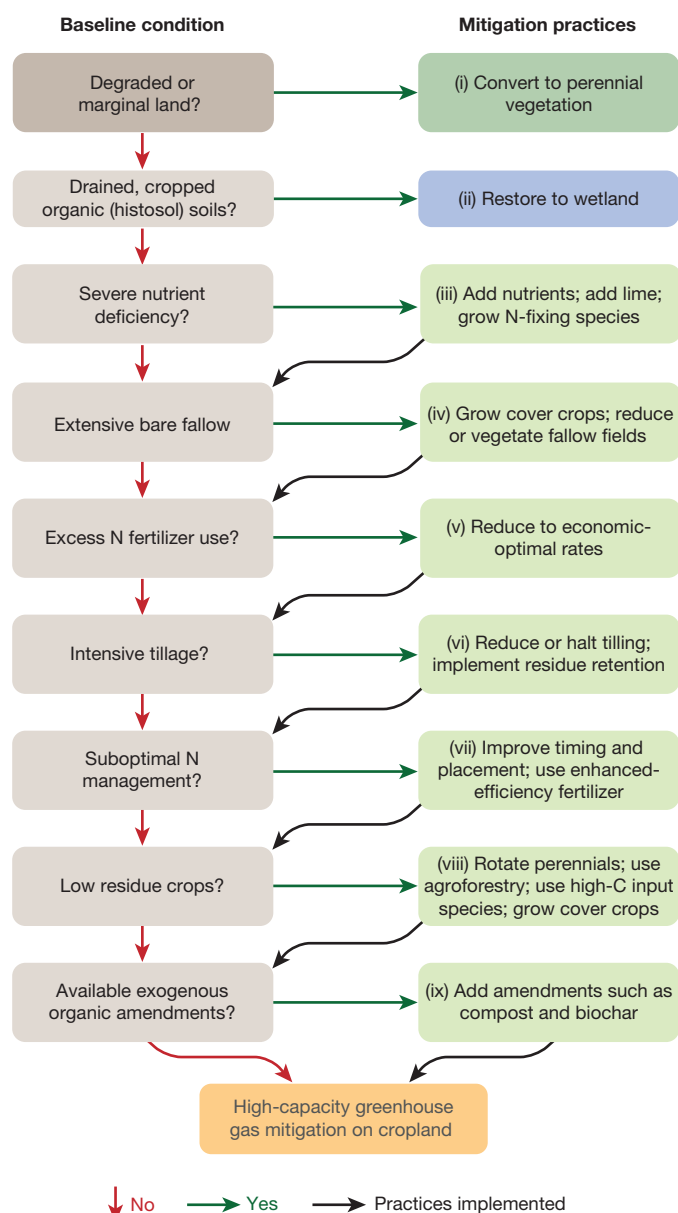


Figure 1 | Decision tree for cropland GHG mitigating practices.

(Rice is not included.) For degraded, marginal lands, the most productive mitigation option is conversion to perennial vegetation either left unmanaged or sustainably harvested to offset fossil energy use (cellulosic biofuels). Histosol is soil with very high organic matter content, such as from peat bog. For more arable lands, multiple options could be implemented sequentially or in combination, depending on management objectives, cost and other constraints (see Table 1). The practices shown (see Table 1 and text for more discussion) are roughly ordered from lower-cost or higher-feasibility options towards more costly interventions (bottom of figure).

In general, soil C sequestration rates on land maintained in agricultural use are lower than for land restoration to grassland or forest, and vary on the order of $0.1\text{--}1\text{ Mg C ha}^{-1}\text{ yr}^{-1}$, as a function of land-use history, soil or climate conditions, and the combination of management practices applied^{2,15}. Practices that increase C inputs include (1) improved varieties or species with greater root mass to deposit C in deeper layers where turnover is slower²², (2) adopting crop rotations that provide greater C inputs²³, (3) more residue retention²⁴, and (4) cover crops during fallow periods to provide year-round C inputs (Fig. 1)^{23,25}. Cover crops can also reduce nutrient losses, including nitrate that is otherwise converted to N_2O in riparian areas and waterways²⁶—an example

of synergy between practices that sequester C and also tighten the N cycle to limit emissions of N_2O (Table 1). Other practices to increase C inputs include irrigation in water-limited systems¹⁹ and additional fertilizer input to increase productivity in low-yielding, nutrient-deficient systems (Fig. 1)²⁷. Although additional nutrient and water inputs to boost yields may increase non- CO_2 emissions²⁸, the emissions intensity of the system (GHG emissions per unit yield) may decline, providing a global benefit if the yield increase avoids land conversion for agriculture elsewhere^{21,23}.

Many croplands can sequester C through less intensive tillage, particularly zero tillage¹⁵, owing to less disruption of soil aggregate structure²⁹. Some authors have argued that benefits are small because increased C content in surface horizons are offset by C losses deeper in the profile³⁰, although others have noted that the larger variability in sub-surface horizons and lack of statistical power in existing studies makes such conclusions questionable³¹.

A change from annual to perennial crops typically increases below-ground C inputs (and soil disturbance is reduced), leading to C sequestration¹⁶. In grasslands, soil C sequestration can be increased through optimal stocking/grazing density³². Improved management in fire-prone ecosystems via fire prevention or prescribed burning can also increase C sequestration³³.

Key knowledge gaps that affect our understanding of soil C sequestration processes and management options to implement them include questions about the differential temperature sensitivities of C turnover among soil organic matter fractions³⁴, interactions among organic matter chemistry, mineral surface interactions and C saturation^{35–37}, and subsoil ($>30\text{ cm}$) soil organic matter accretion, turnover and stabilization³⁸. Landscape processes, particularly the impact of erosion and lateral transport of C in sediments, contribute additional uncertainty on net sequestration occurring at a specific location³⁹. And emerging evidence that stabilized soil organic matter is of microbial rather than direct plant origin^{35,40} may offer a potential to manipulate the soil–plant microbiome to enhance C sequestration in the rhizosphere.

Soil C sequestration via exogenous C inputs

Addition of plant-derived C from external (that is, offsite) sources such as composts or biochar can increase soil C stocks, and may result in net CO_2 removals from the atmosphere (Fig. 1). Both compost and biochar are more slowly decomposed compared to fresh plant residues, with composts typically having mean residence times several times greater than un-composted organic matter⁴¹, and biochar mineralizes 10–100 times more slowly than uncharred biomass⁴². Thus a large fraction of added C—particularly for biochar—can be retained in the soil over several decades or longer, although residence times vary depending on the amendment type, nutrient content and soil conditions³⁶ (such as moisture, temperature and texture).

However, because the organic matter originates from outside the ecosystem ‘boundary’, a broader life-cycle assessment approach is needed, that considers the GHG impacts of: (1) offsite biomass removal, transport, and processing; (2) alternative end uses of the biomass; (3) interactions with other soil GHG-producing processes; and (4) synergies between these soil amendments and the fixation and retention of *in situ* plant-derived C^{43,44}. In many cases, net life-cycle emissions will largely depend on whether the biomass used as a soil amendment would have otherwise been burnt (either for fuel, thereby offsetting fossil fuel use, or as waste disposal), added to a landfill, or left in place as living biomass or detritus^{43,44}.

While slower mineralization of the amendment is an important determinant of net mitigation impact, effects on other soil emissions cannot be neglected. Mineralization of existing soil C in response to amendments (often referred to as ‘priming’⁴⁵) has often been observed immediately following biochar addition, but priming usually declines, sometimes becoming negative (that is, inhibiting *in situ* soil C decomposition), over time^{46,47}. Analogous time dependence of soil N_2O and CH_4 emissions has not received sufficient attention⁴¹. Increased plant growth

Table 1 | Co-benefits, relative costs and constraints for the mitigation practices shown in Fig. 1

Mitigation practices	Practice co-benefits	Relative cost		Constraints and caveats
		Developed	Less developed	
(i)	↓ Soil erosion ↑ Biodiversity ↑ Water quality	\$\$	\$\$	Alternate land/livelihood for subsistence farmers; opportunity cost of removing land base; potential for leakage (that is, land-use change impact)
(ii)	↑ Biodiversity ↑ Water quality	\$\$\$	\$\$\$	High opportunity cost of lost crop production; potential increase in CH ₄ emissions; potential for leakage (that is, land-use change impact)
(iii)	↑ Food security ↑ Water quality	\$	\$\$	Availability or access to fertilizer; potential increase in N ₂ O emissions
(iv)	↓ Soil erosion ↑ Water quality ↑ Soil health ↑ Food security	\$	\$\$	Limited applicability in dry areas
(v)	↑ Water quality	\$	\$	Risk of crop production loss
(vi)	↓ Soil erosion ↑ Water quality ↑ Soil health	\$	\$\$	Limited applicability in cold climates; potential increased equipment cost; increased herbicide use
(vii)	↑ Water quality	\$\$	\$\$\$	Availability or access to enhanced efficiency fertilizer
(viii)	↑ Biodiversity ↑ Water quality ↑ Soil health	\$\$\$	\$\$	Less applicability in dry areas and shallow soils; potential opportunity costs of lost crop production
(ix)	↑ Soil health ↑ Food security	\$\$\$	\$\$	Dependent on life-cycle emissions of producing the amendment

Mitigation practices are numbered according to Fig. 1. Co-benefits include non-GHG ecosystem services from practice implementation. Relative costs are provided as examples based on a developed region such as North America and a less developed region such as sub-Saharan Africa; however, a specific option in one region may have a higher cost or be a less feasible option in another region. Potential constraints include factors that might limit or preclude practice adoption or increase other GHG emissions as a consequence of practice adoption. All options require a region-specific full-cost carbon accounting (GHG life-cycle analysis) that considers potential indirect land-use effects in order to define specific mitigation potentials.

in amended soils and the resultant feedbacks to soil C can make up a large proportion of the soil-based GHG balance^{41,48} and these feedbacks may be especially important for more persistent amendments, because of the longer duration of any effects.

Soil management to reduce N₂O emissions

Arable soils emit more N₂O to the atmosphere than any other anthropogenic source^{2,19}; some 4.2 teragrams (Tg) of a global anthropogenic flux of 8.1 Tg N₂O-N yr⁻¹. Reducing this flux represents a substantial mitigation opportunity, particularly since N₂O is often the major source of radiative forcing in intensively managed cropland. Better N management to reduce emissions would also ameliorate other environmental problems such as nitrate pollution of ground and surface waters caused by excess reactive N in agroecosystems (Fig. 1, Table 1).

N₂O is produced in soils by microbial activity—mainly nitrification and denitrification—which occurs readily when stimulated by the abundant N that cycles rapidly in virtually all agroecosystems. During nitrification, ammonium added as fertilizer, fixed from the atmosphere by legumes, or mineralized from soil organic matter, crop residue, or other inputs is oxidized to nitrite and eventually to nitrate in a series of reactions that can also produce N₂O. Likewise, when denitrifiers use nitrate as an electron acceptor when soil oxygen is low, N₂O is an intermediate product that can readily escape to the atmosphere.

Arable soils managed to support high crop productivity have the capacity to produce large quantities of N₂O, and fluxes are directly related to N inputs. On average, about 1% of the N applied to cropland is directly emitted as N₂O (ref. 49), which is the basis for estimating emissions using default Intergovernmental Panel on Climate Change methods¹⁸. However, recent evidence suggests that this value is too high for crops that are under-fertilized and too low for crops that are fertilized liberally²⁸. When crops compete with microbes for available N, N₂O fluxes are lower. In addition to direct in-field emissions, high N applications cause N losses from leaching and volatilization that contribute to 'indirect' N₂O emissions, downstream or downwind from the field⁵⁰.

Since N₂O has no significant terrestrial sink, abatement is best achieved by attenuating known sources of N₂O emissions, by altering the

environmental factors that affect N₂O production (soil N, oxygen, and C) or by biochemically inhibiting conversion pathways using soil additives. For example, nitrification can be inhibited with commercial additives such as nitrapyrin and dicyandiamide, which slow ammonium oxidation, and field experiments suggest that inhibitors can reduce N₂O fluxes by up to 40% in some soils, although other soils show little reduction and more research is needed to understand variable site-level responses⁵¹. Likewise, tillage and water management can affect N₂O fluxes by altering the soil microenvironment^{52,53}.

Another means of reducing N₂O emissions from arable soils is more precise N management to minimize excess N not used by the crop, while maintaining sustainable high yields. Fertilized crops typically take up less than 50% of the N applied; the remainder is available for loss. According to one recent study⁵⁴, corn farmers in the midwestern states of the USA could reduce N₂O loss by 50% with more conservative fertilizer practices. Nitrogen conservation can be achieved by: (1) better matching application rates of N to crop needs using advanced statistical and quantitative modelling; (2) applying fertilizer at variable rates across a field based on natural patterns of soil fertility, or within the root zone rather than broadcast on the soil surface; and (3) applying fertilizer close to when the crop can use it, such as several weeks after planting, or adding it earlier but using slow-release coatings to delay its dissolution (Fig. 1, Table 1)⁵⁰.

High temporal and spatial variability make predictions of changes in N₂O fluxes in response to management surprisingly difficult. Particularly lacking are empirical data for multi-intervention strategies that may interact in unexpected ways. Aligned to this paucity are gaps in our understanding of how N cycling and net N₂O flux in managed soils will respond to future climate change⁵⁵. The limited number of field manipulation studies to date indicate that changing temperature and precipitation patterns may have large and strongly coupled effects on net N₂O emissions⁵⁶, yet our understanding of the processes that underpin these effects and their robust representation in models is far from complete.

Soil management to reduce CH₄ emissions

More than one-third (>200 Tg yr⁻¹)⁹ of global methane (CH₄) emissions occur through the microbial breakdown of organic compounds in

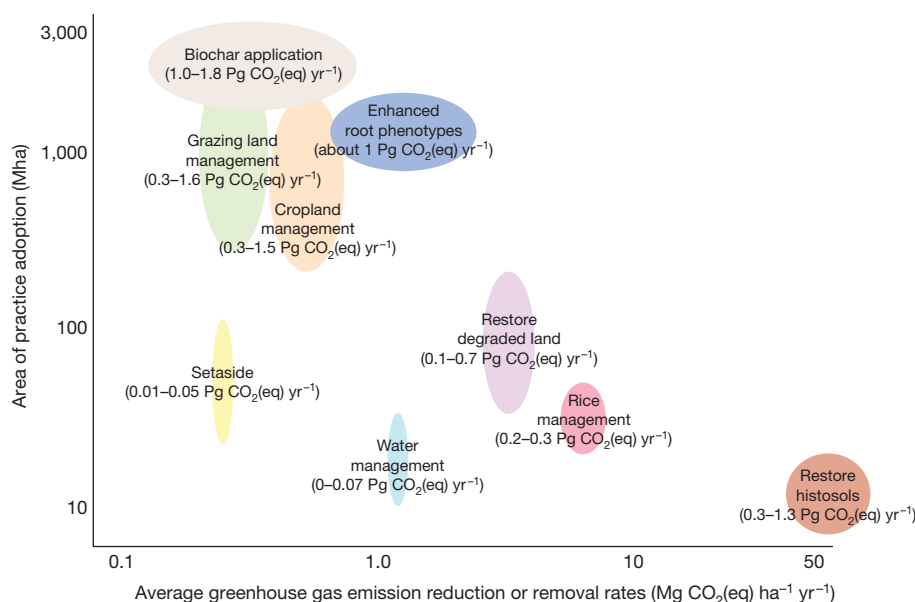


Figure 2 | Global potential for agricultural-based GHG mitigation practices. Management categories are arranged according to average per hectare net GHG reduction rates and potential area (in millions of hectares) of adoption (note log-scales). Unless otherwise noted, estimates are from ref. 19, based on cropland and grassland area projections for 2030. Ranges given in units of total Pg CO₂(eq) yr⁻¹ represent varying adoption rates as a function of C pricing (US\$20, US\$50 and US\$100 per Mg CO₂(eq)), to a maximum technical potential—that is, the full implementation of practices on the available land base. Multiple practices are aggregated for cropland (for example, improved crop rotations and nutrient management, reduced tillage) and grazing land (for example, grazing management, nutrient and fire management, species introduction) categories. Practices that increase net soil C stocks or reduce emissions of N₂O and CH₄ are combined in each practice category.

soils under anaerobic conditions⁵⁷. As such, wetlands (177–284 Tg yr⁻¹) and rice cultivation (33–40 Tg yr⁻¹)⁹ represent the largest soil-mediated sources of CH₄ globally. In contrast, well aerated soils act as sinks for CH₄ (estimated at ~30 Tg yr⁻¹) from the atmosphere via CH₄ oxidation, the bulk of this net sink being in unmanaged upland and forest soils⁵⁸.

Key determinants of soil CH₄ fluxes include aeration, substrate availability, temperature and N inputs⁵⁹; therefore, soil management can radically alter CH₄ fluxes. For example, in most soils, conversion to agriculture severely restricts CH₄ oxidation, related to the suppression of methanotrophs by accelerated N cycling⁶⁰. In flooded rice, alterations in drainage regimes and organic residue incorporation could reduce emissions by ~25% or 7.6 Tg CH₄ yr⁻¹ globally¹⁹, although cycles of wetting and drying of soils may also enhance N₂O production⁶¹ and soil C mineralization⁶², thereby reducing the net mitigation effect.

With global rice production projected to expand by ~40% between 2000 and 2023 (ref. 63), the potential for further GHG mitigation via soil management appears to be large, although the global distribution and diverse nature of rice production systems—including irrigated, rain-fed and deep-water—present challenges to developing effective mitigation strategies. For longer-term (>20 year) projections, climate change and land–atmosphere interactions become increasingly important, with changes in N inputs, temperature, precipitation and atmospheric CO₂ concentration all likely to affect net CH₄ fluxes from soils⁶⁴.

This uncertainty highlights important gaps in understanding key processes and their underlying controls. The restoration of soil CH₄ uptake following agricultural conversion, for example, appears related to methanotroph community diversity⁶⁵, about which we know too little. Likewise the abatement of CH₄ generation in rice rhizospheres is related to C compounds exuded by roots, such that CH₄ mitigation might be achieved through further rice breeding and genetics⁶⁶.

The portion of projected mitigation from soil C stock increase (about 90% of the total technical potential) would have a limited time span of 20–30 years, whereas non-CO₂ emission reduction could, in principle, continue indefinitely¹⁹. Estimates for biochar application⁶⁷ represent a technical potential only, but it is based on a full life-cycle analysis applicable over a 100-year time span. Although global estimates of the potential impact of enhanced root phenotypes for crops have not been published, a first-order estimate of about 1 Pg CO₂(eq) yr⁻¹ is shown, using the global average C accrual rates (0.23 Mg C ha⁻¹ yr⁻¹) for cover crops²⁵, applied to 50% of the cropland land area used by ref. 19. ‘Setaside’ land is arable land, usually for annual crops, that is taken out of production and converted to perennial vegetation (often grassland) and not actively managed for agricultural production, such as conservation reserves.

Limited availability of field-scale CH₄ flux data means a greater reliance on regionally averaged emission factors and extrapolation from mesocosm and laboratory incubations¹⁸, and thus less site and condition specificity in modelling fluxes. Importantly, establishing the net climate forcing effects of any intervention is a prime target for future soil management research.

Global potential for soil GHG mitigation

How important, in total, is this large, varied set of land-use and management practices as a GHG mitigation strategy? One of the challenges in answering this question is to distinguish between what is technically feasible and what might be achieved given economic, social and policy constraints. A comprehensive global analysis of agricultural practices¹⁹ combined climate-stratified modelling of emission reductions and soil C sequestration with economic and land-use change models to estimate mitigation potential as a function of varying ‘C prices’ (reflecting a social incentive to pay for mitigation). They estimated total soil GHG mitigation potential ranging from 5.3 Pg CO₂(eq) yr⁻¹ (without economic constraints) to 1.5 Pg CO₂(eq) yr⁻¹ at the lowest specified C price (US\$20 per Mg of CO₂(eq)). Average rates for the majority of management interventions are modest, <1 Mg CO₂(eq) ha⁻¹ yr⁻¹. Thus, achieving large global GHG reductions requires a substantial proportion of the agricultural land base (Fig. 2). Although the economic and management constraints on biochar additions (not assessed by ref. 19) are less well known, ref. 67 estimated a global technical potential of 1–1.8 Pg CO₂(eq) yr⁻¹ (Fig. 2).

A more unconventional intervention that has been proposed is the development of crops with larger, deeper root systems, hence increasing plant C inputs and soil C sinks^{22,68}. Increasing root biomass and selecting for root architectures that store more C in soils has not previously

been an objective for crop breeders, although most crops have sufficient genetic plasticity to alter root characteristics substantially⁶⁹ and selection aimed at improved root adaptation to soil acidity, hypoxia and nutrient limitations could yield greater root C inputs as well as increased crop yields⁶⁸. Greater root C input is well recognized as a main reason for the higher soil C stocks maintained under perennial grasses compared to annual crops¹⁶. Although there are no published estimates of the global C sink potential for 'root enhancement' of annual crop species, as a first-order estimate, a sustained increase in root C inputs might add $\sim 1 \text{ Pg CO}_2(\text{eq}) \text{ yr}^{-1}$ or more if applied over a large portion of global cropland area (Fig. 2).

Thus, the overall mitigation potential of existing (and potential future) soil management practices could be as high as $\sim 8 \text{ Pg CO}_2(\text{eq}) \text{ yr}^{-1}$. How much is achievable will depend heavily on the effectiveness of implementation strategies and socioeconomic and policy constraints. A key strength is that a variety of practices can often be implemented on the same land area, to leverage synergies, while avoiding offsetting effects for different gases (Fig. 1). But regardless of which combination of management interventions are pursued, effective policies, that incentivize land managers to adopt them, will be needed. A common thread across implementation strategies is the role for strong science-based metrics with which to measure and monitor performance.

Implementation of mitigation practices

Relative to many other GHG source categories, agricultural soil GHG mitigation presents particular challenges. Rates on an individual land parcel are often low, but vast areas of land are devoted to agriculture globally, and the implementers of mitigation practices—the people using the land—number in the billions. Therefore, engaging a substantial number of these people is a massive undertaking in itself. Furthermore, agricultural soil GHG emissions are challenging to quantify owing to their dispersed and variable nature and the multiplicity of controlling factors—operating across heterogeneous landscapes. Direct measurement of fluxes requires specialized personnel and equipment, normally limited to research environments, and hence not feasible for most mitigation projects. Model-based methods, in which emission rates are quantified as a function of location, environmental conditions and management, provide a more feasible approach^{54,70,71}. Process-based models, which dynamically simulate mechanisms and controls on fluxes as a function of climatic and soil variables and management practices, and empirical models based on statistical analysis of field-measured flux rates, represent differing but complementary approaches. In general, model-based quantification systems enable monitoring to focus on practice performance and thus dramatically reduce transaction costs for implementing mitigation policies⁷⁰.

Several implementation strategies for soil GHG mitigation exist (see Box 1), all of which require robust quantification and monitoring technologies. Those requiring the most rigorous methods involve offset projects participating in cap-and-trade markets, in which land managers are directly compensated for achieving emission reductions. Other market-linked strategies, such as 'green labelling' systems for agricultural products, will also require rigorous yet easy-to-use GHG quantification tools, enabling agricultural producers to meet standards set by product distributors and accepted by consumers^{7,72}.

Within the voluntary C offset market space, there are a growing number of projects that include soil GHG mitigation components⁵. Several large projects focus on preventing land conversion (that is, from forest and grassland), thus avoiding large CO_2 emissions from soils and liquidated biomass C stocks. Relatively simple empirical models supplemented with field measurements are commonly used for avoided land conversion projects. For more complex land-use projects, empirical models are less suited to capturing interactions across multiple emission sources, and may over- or under-credit projects where a practice has an influence on multiple emission sources. There are relatively fewer projects targeting GHG mitigation on existing agricultural lands, involving a broader suite of soil management practices, and early pilot-phase

BOX 1

Implementing soil GHG reductions

Incentives for farmers to adopt alternative practices that mitigate GHGs can take a variety of forms, including:

(1) Regulation and taxation. Direct regulatory measures intended to reduce soil GHGs at the farm scale are probably politically unfeasible and costly. Taxation of N fertilizer, already used in parts of the USA and Europe to reduce nitrate pollution, could function as an indirect tax that would reduce N_2O emissions.

(2) Subsidies. Targeted government payments or subsidies for implementing GHG-reducing practices is emerging as a policy alternative. For example, US Department of Agriculture programmes are including GHG mitigation as a conservation goal and provisions in the EU Common Agricultural Policy link subsidy payments to 'cross-compliance' measures that include maintenance of soil organic matter stocks⁹⁹. A more direct link to soil GHG emissions follows from a recent decision to include cropland and grassland in EU commitments under the Kyoto Protocol¹⁰⁰.

(3) Supply-chain initiatives. Major food distributors are targeting sustainability metrics, including low GHG footprints, as a consumer marketing strategy (see refs 101 and 102 for examples of initial efforts involving agricultural producers and product distributors and retailers), setting performance standards for contracted agricultural producers, including the requirement of field-scale monitoring of production practices and quantification of GHG emissions.

(4) Cap and trade. In a cap-and-trade system, emitters are subject to an overall emissions level or 'cap', in which permitted emissions decrease over time. Emitters can stay below the capped levels by reducing their own emissions or by purchasing surplus permits from capped entities that have exceeded their required reductions. Both compliance and voluntary markets can function as cap-and-trade systems¹⁰³. Within many cap-and-trade systems, a limited amount of emission reductions (termed 'offsets') can be provided by non-capped entities. The inclusion of agricultural activities as offset providers has been growing, particularly within voluntary markets. To maintain the integrity of emission caps, key criteria for offset providers include demonstrating additionality (that is, ensuring that reductions result from project interventions and not simply business-as-usual trends), avoiding leakage (that is, unintended emission increases elsewhere as a consequence of the project activities), and providing for permanence (meaning that increased soil C storage, credited as a CO_2 removal, is maintained long-term).

N_2O and CH_4 reduction projects are only now being developed^{5,54}. Here, accurately quantifying C sequestration and/or emission reductions is more challenging owing to lower rates of change relative to baseline conditions, thus requiring more sophisticated models and supporting research infrastructure (Fig. 3).

Another challenge for projects on existing agricultural lands is obtaining and processing the management activity data. For example, the Kenya Agriculture Carbon Project (KACP) involves a total of 60,000 individual small-holder farmers⁷³. In contrast to projects involving major land-cover changes, where remote sensing can provide much of the activity monitoring (for example, retention of forested land over time), remote-sensing options are poorly suited for monitoring crop type, fertilizer, residue and water management, and organic matter amendments⁷⁴; for such practices the best sources of information are the land managers themselves (Fig. 3).

Thus another option is to engage land managers as information providers. Examples of this approach are the Cool Farm Tool (<http://www.coolfarmtool.org/>)⁷², being used by farmers participating in low C supply chain management, and the COMET-Farm tool

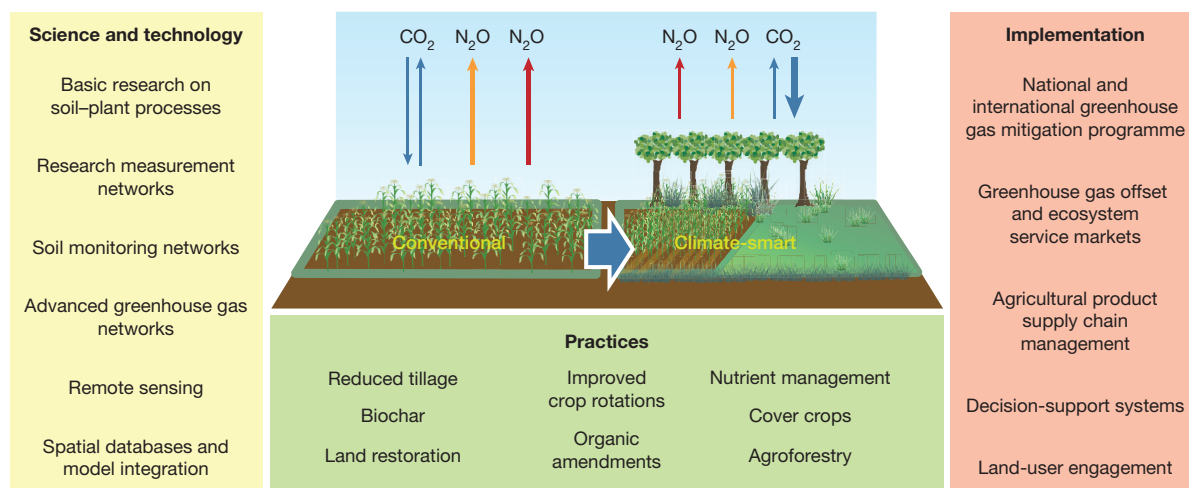


Figure 3 | Expanding the role of agricultural soil GHG mitigation will require an integrated research support and implementation platform.

Targeted basic research on soil processes, expanding measurement and monitoring networks, and further developing global geospatial soils data can improve predictive models and reduce uncertainties. Ongoing advances in information technology and complex system and 'Big Data' integration offer the potential to engage a broad-range of stakeholders, including land managers, to 'crowd-source' local knowledge of agricultural

management practices through web-based computer and mobile apps, and help drive advanced model-based GHG metrics. This will facilitate the implementation of climate-smart soil management policies, via cap-and-trade systems, product supply-chain initiatives for 'low-carbon' consumer products, and national and international GHG mitigation policies; it will also promote more sustainable and climate-resilient agricultural systems, globally.

(<http://cometfarm.nrel.colostate.edu>), which allows farmers to compute full farm-scale GHG budgets, for support of government-sponsored conservation initiatives and participation in mitigation projects⁷⁵. Both tools provide web-based interfaces designed for non-specialists to enter land management information; Cool Farm utilizes empirical emission factor-type models, while COMET-Farm incorporates both empirical and process-based models. Such systems can be used to integrate local knowledge on management practices with detailed soil and climate maps, remote sensing and sophisticated models for emission calculations. Soon much of this functionality could be deployed in mobile applications (Fig. 3), which would be particularly advantageous in developing countries where the existing infrastructure to collect and manage land-use data is weak⁷⁶.

Quantifying uncertainties

Inventories of soil C stock changes and net GHG fluxes using process-based models will always have uncertainty due to lack of process understanding, inadequate parameterization, and limitations associated with model inputs⁷⁷ (such as weather, management and soils data). Empirical models generally rely on statistical analyses of measurement data to produce emission factors, along with an estimated uncertainty¹⁵. However, empirical models can be biased if measurements do not fully reflect the conditions for the agroecosystems in the project. Even with the limitations in process-based understanding, process-based models are likely to provide the most robust framework for estimating soil C stock and GHG flux changes in climate-smart agriculture programmes⁷⁸.

Monitoring, reporting and verification systems are a key element in a climate-smart agricultural programme. Such systems place different levels of importance on uncertainty depending on program type (see Box 1)⁷⁹, but discounting payments on the basis of the level of uncertainty is likely to be part of programmes with financial incentives, such as cap-and-trade. Discounting encourages monitoring efforts to reduce uncertainty over time¹⁸. If discounting payments for C sequestration and emission reduction practices with larger uncertainty is adopted in climate-smart agricultural programmes, then more advanced methods with process-based models will probably emerge as the preferred method because they will tend to have less uncertainty. For example, uncertainty was reduced by 24% when predicting national-scale C stock changes in the United States with process-based models compared to empirically derived factors⁷⁷.

Another consideration is that uncertainties in estimating C stock and GHG emissions with process-based models are considerably larger for reporting by single individuals, particularly if the amount of change on an individual farm is small⁷⁷. Aggregation of many farms into larger projects will reduce uncertainties, which could be a viable approach for managing uncertainty and reducing the discounting of incentive payments.

Verification is an independent evaluation of estimated emissions intended to provide confidence that the reported results are correct, but in practice, the requirements for verification are highly variable across different GHG mitigation efforts, ranging from essentially no requirements to annual evaluations⁷⁹. Verification typically focuses on the accuracy of the estimates, and possibly the most stringent approach is an independent set of measurements. Although independent data may be less favoured in terms of costs relative to alternatives, such as expert judgement⁷⁹, soil monitoring networks deployed at national or regional scales could produce independent data for evaluating model-based assessments of soil C stock changes and GHG emissions⁸⁰ and for model bias adjustment, using empirically based methods⁸¹.

Another approach to verification is to use atmospheric observations of trace-gas concentrations and inverse modelling to estimate fluxes between the atmosphere and land surface^{82,83}. This 'top-down' modelling, using a network of tower-based observations of CO₂ concentrations, was used to verify 'bottom-up' inventory modelling based on observed management activities in the largely agricultural region of the central USA^{84,85}. Since atmospheric observations integrate all CO₂ fluxes in the region, the inventory included a full assessment of all sources and sinks. However, even with the fully integrated CO₂ flux, it is possible to statistically disaggregate individual sources as part of the analysis, such as contributions from soil C pools to the regional flux⁸⁶. Satellite-based measurements are providing a new source of atmospheric trace-gas data that can be used to estimate land surface fluxes with inverse modelling frameworks^{87,88}. Although atmospheric observations and satellite imagery may become a standard for verifying regional inventories in the future, these methods need further testing before operational systems can be reliably deployed.

Conclusions and recommendations

Climate change and GHG mitigation require an 'all of the above' approach⁸⁹, where all reduction measures that are feasible, cost-effective and environmentally sustainable should be pursued. For soils, a variety

of management practices and technologies are known to reduce emissions and promote C sequestration, most of which also provide environmental co-benefits. The impediment to implementing agricultural soil GHG mitigation strategies more aggressively to date is primarily the feasibility of quantifying and verifying soil mitigation activities⁹⁰ in a cost-effective manner. Overcoming this barrier therefore translates into: (1) increasing the acceptance of soil management within compliance and voluntary C markets; (2) reducing costs to governments for providing environment-based subsidies; and (3) meeting the demands of consumers for 'low carbon' products.

Reducing and managing uncertainties are key to both improving predictive models and decision-support tools and the design of effective policies that promote soil-based GHG mitigation. To advance these efforts, several research and development priorities are apparent (Fig. 3). First, support for research site networks of soil flux (N_2O , CH_4) and soil C measurements encompassing a wide variation in management, as well as 'on-farm' soil C monitoring networks⁸⁰ needs to be strengthened (see ref. 91 for an example of a large-scale research network measuring soil GHG flux in a geographically distributed field experiment, using uniform protocols, advanced instrumentation and data portals in Europe, and ref. 92 for an example in the USA). Such support should coordinate with basic research (for example, on soil organic matter stabilization processes, N_2O and CH_4 microbiology, plant-microbe interactions, plant breeding and root phenotyping) to advance process understanding, develop new mitigation practices and fill gaps in underrepresented soil- and climate-management systems. High-quality data generated from consistent measurement protocols is critical for evaluating and improving models. These efforts may benefit from development of new sensor technologies, enabling cheaper and quicker soil measurements⁹³. Although multiple competing models are needed, both to spur innovation and because no single model will be best in all situations, model development will benefit from greater collaboration and cross-model testing among developers, moving towards a more open-source, community-development approach⁹⁴. Large geo-spatial databases of soil biophysical properties and climate variables are critical to quantify soil processes accurately across the landscape (Fig. 3). High-resolution soil maps exist in most developed countries (and increasingly in developing countries⁹⁵), and if made publicly available (there is free access to fine-spatial-scale (about 1:15,000 to 1:20,000) soil maps for the USA⁹⁶), would greatly improve capabilities for modelling GHG emissions at scale.

Finally, realizing the potential for climate change mitigation through global soil management requires understanding cultural, political and socioeconomic contexts, and the ways in which widespread, sustained changes in practice can be successfully achieved within it^{97,98}. As such, there needs to be a greater level of engagement with the land users themselves, who will be the ones implementing practices that abate GHG emissions and sequester C. Engagement means both education and outreach, highlighting the links between agriculture and GHGs and using innovative strategies⁷⁶ (Fig. 3) to involve stakeholders in gathering and using their local knowledge of how the land is being used now and how it might best be used in the future, thus establishing a new paradigm for climate-smart soil management.

Received 9 February 2015; accepted 27 January 2016.

- Ruddiman, W. F. The anthropogenic greenhouse era began thousands of years ago. *Clim. Change* **61**, 261–293 (2003).
- Smith, P. et al. Agriculture, Forestry and Other Land Use (AFOLU). In *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 813–922 (Cambridge Univ. Press, 2014).
- Tubiello, F. N. et al. The contribution of agriculture, forestry and other land use activities to global warming, 1990–2012. *Glob. Change Biol.* **21**, 2655–2660 (2015).
- Smith, P. Soils and climate change. *Curr. Opin. Environ. Sust.* **4**, 539–544 (2012).
- Verified Carbon Standard (VCS) <http://www.v-c-s.org/> (accessed May 2015).
- American Carbon Registry (ACR) <http://americancarbonregistry.org/> (accessed May 2015).
- Lavallée, S. & Plouffe, S. The ecolabel and sustainable development. *Int. J. Life Cycle Assess.* **9**, 349–354 (2004).
- Kahiluoto, H., Smith, P., Moran, D. & Olesen, J. E. Enabling food security by verifying agricultural carbon sequestration. *Nature Clim. Change* **4**, 309–311 (2014).
- Ciais, P. et al. Carbon and other biogeochemical cycles. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 465–570 (Cambridge Univ. Press, 2013).
- Join the 4% Initiative Soils for Food Security and Climate <http://agriculture.gouv.fr/join-40-initiative-soils-food-security-and-climate-0> (accessed 23 Sept 2015).
- Batjes, N. H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* **47**, 151–163 (1996).
- Paustian, K. et al. Agricultural soil as a C sink to offset CO₂ emissions. *Soil Use Manage.* **13**, 230–244 (1997).
- West, T. O. & Six, J. Considering the influence of sequestration duration and carbon saturation on estimates of soil carbon capacity. *Clim. Change* **80**, 25–41 (2007).
- Davidson, E. A. & Ackerman, I. L. Changes in soil carbon inventories following cultivation of previously untilled soils. *Biogeochemistry* **20**, 161–193 (1993).
- Ogle, S. M., Breidt, F. J. & Paustian, K. Agricultural management impacts on soil organic carbon storage under moist and dry climatic conditions of temperate and tropical regions. *Biogeochemistry* **72**, 87–121 (2005).
- Conant, R. T., Paustian, K. & Elliott, E. T. Grassland management and conversion into grassland: effects on soil carbon. *Ecol. Appl.* **11**, 343–355 (2001).
- Guo, L. B. & Gifford, R. M. Soil carbon stocks and land use change: a meta analysis. *Glob. Change Biol.* **8**, 345–360 (2002).
- IPCC. *Intergovernmental Panel on Climate Change Guidelines for National Greenhouse Gas Inventories* (eds Eggleston, S., Buendia, L., Miwa, K., Ngara, T. & Tanabe, K.) Vol. 4, *Agriculture, Forestry and Other Land Use (AFOLU)* <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html> (Institute for Global Environmental Strategies, 2006).
- Smith, P. et al. Greenhouse gas mitigation in agriculture. *Phil. Trans. R. Soc. B* **363**, 789–813 (2008).
- A comprehensive analysis of agricultural GHG emissions and mitigation potentials including estimated C price impacts on mitigation activities.**
- Knox, S. H. et al. Agricultural peatland restoration: effects of land-use change on greenhouse gas (CO₂ and CH₄) fluxes in the Sacramento-San Joaquin Delta. *Glob. Change Biol.* **21**, 750–765 (2015).
- Foley, J. et al. Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
- Kell, D. Large-scale sequestration of atmospheric carbon via plant roots in natural and agricultural ecosystems: why and how. *Phil. Trans. R. Soc. B* **367**, 1589–1597 (2012).
- Burney, J. A. et al. Greenhouse gas mitigation by agricultural intensification. *Proc. Natl Acad. Sci. USA* **107**, 12052–12057 (2010).
- Wilhelm, W. W., Johnson, J. M. F., Hatfield, J. L., Voorhees, W. B. & Linden, D. R. Crop and soil productivity response to corn residue removal: a literature review. *Agron. J.* **96**, 1–17 (2004).
- Poeplau, C. & Don, A. Carbon sequestration in agricultural soils via cultivation of cover crops—a meta-analysis. *Agric. Ecosyst. Environ.* **200**, 33–41 (2015).
- This paper combines an analysis of globally distributed field data with simulation modelling to quantify potential soil C increases with adoption of cover crops on previously fallow soils.**
- Tonitto, C., David, M. B. & Drinkwater, L. E. Replacing bare fallows with cover crops in fertilizer-intensive cropping systems: a meta-analysis of crop yield and N dynamics. *Agric. Ecosyst. Environ.* **112**, 58–72 (2006).
- Lemke, R. L., VandenBygaart, A. J., Campbell, C. A., Lafond, G. P. & Grant, B. Crop residue removal and fertilizer N: effects on soil organic carbon in a long-term crop rotation experiment on a Udic Boroll. *Agric. Ecosyst. Environ.* **135**, 42–51 (2010).
- Shcherbak, I., Millar, N. & Robertson, G. P. Global meta-analysis of the nonlinear response of soil nitrous oxide (N₂O) emissions to fertilizer nitrogen. *Proc. Natl Acad. Sci. USA* **111**, 9199–9204 (2014).
- This paper shows that N₂O emissions are greater than previously thought for soils receiving high rates of N fertilizer.**
- Six, J., Elliott, E. T. & Paustian, K. Soil macroaggregate turnover and microaggregate formation: a mechanism for C sequestration under no-tillage agriculture. *Soil Biol. Biochem.* **32**, 2099–2103 (2000).
- Powlson, D. S. et al. Limited potential of no-till agriculture for climate change mitigation. *Nature Clim. Change* **4**, 678–683 (2014).
- Kravchenko, A. N. & Robertson, G. P. Whole-profile soil carbon stocks: the danger of assuming too much from analyses of too little. *Soil Sci. Soc. Am. J.* **75**, 235–240 (2011).
- McSherry, M. E. & Ritchie, M. E. Effects of grazing on grassland soil carbon: a global review. *Glob. Change Biol.* **19**, 1347–1357 (2013).
- Scholes, M. J. & Scholes, R. J. Dust unto dust. *Science* **342**, 565–566 (2013).
- Conant, R. T. et al. Temperature and soil organic matter decomposition rates—synthesis of current knowledge and a way forward. *Glob. Change Biol.* **17**, 3392–3404 (2011).

35. Cotrufo, M. F., Wallenstein, M., Boot, C. M., Denef, K. & Paul, E. A. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? *Glob. Change Biol.* **19**, 988–995 (2013).
36. Schmidt, M. W. I. *et al.* Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).
37. Stewart, C. E., Plant, A. F., Paustian, K., Conant, R. & Six, J. Soil carbon saturation: linking concept and measurable carbon pools. *Soil Sci. Soc. Am. J.* **72**, 379–392 (2008).
38. Rumpel, C. & Koegel-Knabner, I. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. *Plant Soil* **338**, 143–158 (2011).
39. Nadeu, E., Gobin, A., Fiener, P., van Wesemael, B. & van Oost, K. Modelling the impact of agricultural management on soil carbon stocks at the regional scale: the role of lateral fluxes. *Glob. Change Biol.* **21**, 3181–3192 (2015).
40. Grandy, A. S. & Neff, J. C. Molecular C dynamics downstream: the biochemical decomposition sequence and its impact on soil organic matter structure and function. *Sci. Total Environ.* **404**, 297–307 (2008).
41. Ryals, R., Hartman, M. D., Parton, W. J., DeLonge, M. & Silver, W. L. Long-term climate change mitigation potential with organic matter management on grasslands. *Ecol. Appl.* **25**, 531–545 (2015).
- Field experiments and modelling were used to show reductions in net GHG emissions from grazed annual grasslands using composted green waste applications.**
42. Lehmann, J. *et al.* In *Biochar for Environmental Management: Science, Technology and Implementation* (eds Lehmann, J. & Joseph, S.) 235–282 (Taylor and Francis, 2015).
43. Roberts, K., Gloy, B., Joseph, S., Scott, N. & Lehmann, J. Life cycle assessment of biochar systems: estimating the energetic, economic and climate change potential. *Environ. Sci. Technol.* **44**, 827–833 (2010).
44. DeLonge, M. S., Ryals, R. & Silver, W. L. A lifecycle model to evaluate carbon sequestration potential and greenhouse gas dynamics of managed grasslands. *Ecosystems* **16**, 962–979 (2013).
45. Kuzyakov, Y. Priming effects: interactions between living and dead organic matter. *Soil Biol. Biochem.* **42**, 1363–1371 (2010).
46. Zimmerman, A., Gao, B. & Ahn, M. Y. Positive and negative mineralization priming effects among a variety of biochar-amended soils. *Soil Biol. Biochem.* **43**, 1169–1179 (2011).
47. Whitman, T., Zhu, Z. & Lehmann, J. Carbon mineralizability determines interactive effects on mineralization of pyrogenic organic matter and soil organic carbon. *Environ. Sci. Technol.* **48**, 13727–13734 (2014).
48. Whitman, T., Nicholson, C. F., Torres, D. & Lehmann, J. Climate change impact of biochar cook stoves in Western Kenyan farm households: system dynamics model analysis. *Environ. Sci. Technol.* **45**, 3687–3694 (2011).
49. Bouwman, A. F., Boumans, L. J. M. & Batjes, N. H. Emissions of N₂O and NO from fertilized fields: summary of available measurement data. *Glob. Biogeochem. Cycles* **16**, 1058 (2002).
50. Robertson, G. P. & Vitousek, P. M. Nitrogen in agriculture: balancing the cost of an essential resource. *Annu. Rev. Environ. Resour.* **34**, 97–125 (2009).
51. Akiyama, H., Yan, X. Y. & Yagi, K. Evaluation of effectiveness of enhanced-efficiency fertilizers as mitigation options for N₂O and NO emissions from agricultural soils: meta-analysis. *Glob. Change Biol.* **16**, 1837–1846 (2010).
52. Aguilera, E., Lassaletta, L., Sanz Cobeña, A., Garnier, J. & Vallejo Garcia, A. The potential of organic fertilizers and water management to reduce N₂O emissions in Mediterranean climate cropping systems. *Agric. Ecosyst. Environ.* **164**, 32–52 (2013).
53. van Kessel, C. *et al.* Climate, duration, and N placement determine N₂O emissions in reduced tillage systems: a meta-analysis. *Glob. Change Biol.* **19**, 33–44 (2013).
54. Millar, N., Robertson, G. P., Grace, P. R., Gehl, R. J. & Hoben, J. P. Nitrogen fertilizer management for nitrous oxide (N₂O) mitigation in intensive corn (maize) production: an emissions reduction protocol for US Midwest agriculture. *Mitig. Adapt. Strat. Glob. Change* **15**, 185–204 (2010).
55. Robertson, G. P. *et al.* Nitrogen–climate interactions in US agriculture. *Biogeochemistry* **114**, 41–70 (2013).
56. Reay, D. S. *et al.* Global agriculture and nitrous oxide emissions. *Nature Clim. Change* **2**, 410–416 (2012).
57. Conrad, R. The global methane cycle: recent advances in understanding the microbial processes involved. *Environ. Microbiol. Rep.* **1**, 285–292 (2009).
58. Le Mer, J. & Roger, P. Production, oxidation, emission and consumption of methane by soils. *Eur. J. Soil Biol.* **37**, 25–50 (2001).
59. Segers, R. Methane production and methane consumption: a review of processes underlying wetland methane fluxes. *Biogeochemistry* **41**, 23–51 (1998).
60. Suwanwaree, P. & Robertson, G. P. Methane oxidation in forest, successional, and no-till agricultural ecosystems: effects of nitrogen and soil disturbance. *Soil Sci. Soc. Am. J.* **69**, 1722–1729 (2005).
61. Linquist, B. A. *et al.* Reducing greenhouse gas emissions, water use, and grain arsenic levels in rice systems. *Glob. Change Biol.* **21**, 407–417 (2015).
62. Liu, Y. *et al.* Carbon dioxide flux from rice paddy soils in central China: effects of intermittent flooding and draining cycles. *PLoS ONE* **8**, e56562 (2013).
63. Organisation for Economic Co-operation and Development (OECD)–Food and Agriculture Organization (FAO) of the United Nations. *Agricultural Outlook 2015–2024*. <http://stats.oecd.org> (worldwide rice production for 2000 to 2024 period, accessed March 2016).
64. van Groenigen, K. J., van Kessel, C. & Hungate, B. A. Increased greenhouse-gas intensity of rice production under future atmospheric conditions. *Nature Clim. Change* **3**, 288–291 (2012).
65. Levine, U. *et al.* Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *ISME J.* **5**, 1683–1691 (2011).
66. Su, J. *et al.* Expression of barley SUSIBA2 transcription factor yields high-starch low-methane rice. *Nature* **523**, 602–606 (2015).
67. Woolf, D., Amonette, J. E., Street-Perrott, F. A., Lehmann, J. & Joseph, S. Sustainable biochar to mitigate global climate change. *Nature Commun.* **1**, 56 (2010).
- This paper calculates the global technical potential of greenhouse gas emission reductions by biochar systems as a function of its effects on soil improvement.**
68. Lynch, J. P. & Wojciechowski, T. Opportunities and challenges in the subsoil: pathways to deeper rooted crops. *J. Exp. Bot.* **66**, 2199–2210 (2015).
- This paper looks at the potential for breeding plants with more root production and tolerance to 'problem soils', to increase yields, nutrient and water capture, and C sequestration.**
69. Smith, S. & De Smet, I. Root system architecture: insights from *Arabidopsis* and cereal crops introduction. *Phil. Trans. R. Soc. B* **367**, 1441–1452 (2012).
70. Conant, R. T., Ogle, S. M., Paul, E. A. & Paustian, K. Measuring and monitoring soil organic carbon stocks in agricultural lands for climate mitigation. *Front. Ecol. Environ.* **9**, 169–173 (2011).
71. Paustian, K., Ogle, S. M. & Conant, R. T. In *Handbook of Climate Change and Agroecosystems: Impact, Adaptation and Mitigation* (eds Hillel, D. & Rosenzweig, C.) 307–341 (Imperial College Press, 2011).
72. Hillier, J. G. A farm-focused calculator for emissions from crop and livestock production. *Environ. Modell. Softw.* **26**, 1070–1078 (2011).
73. Swallow, B. M. & Goddard, T. W. Value chains for bio-carbon sequestration services: lessons from contrasting cases in Canada, Kenya and Mozambique. *Land Use Policy* **31**, 81–89 (2013).
74. National Academy of Sciences. *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements* http://www.nap.edu/download.php?record_id=12883 (eds Pacala, S. *et al.*) 1–110 (National Academies Press, 2010).
75. Paustian, K. *et al.* in *Managing Agricultural Greenhouse Gases: Coordinated Agricultural Research through GraceNet to Address Our Changing Climate* (eds Liebig, M., Franzluebbers, A. & Follett, R.) 251–270 (Academic Press, 2012).
76. Paustian, K. Bridging the data gap: engaging developing country farmers in greenhouse gas accounting. *Environ. Res. Lett.* **8**, 021001 (2013).
- This paper proposes the development of mobile apps to 'crowd-source' local-scale knowledge of land use and management data to help improve GHG inventories and project-scale accounting in developing countries.**
77. Ogle, S. M. *et al.* Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. *Glob. Change Biol.* **16**, 810–822 (2010).
78. Smith, P. *et al.* Towards an integrated global framework to assess the impacts of land use and management change on soil carbon: current capability and future vision. *Glob. Change Biol.* **18**, 2089–2101 (2012).
79. Bellassen, V. *et al.* Monitoring, reporting and verifying emissions in the climate economy. *Nature Clim. Change* **5**, 319–328 (2015).
80. van Wesemael, B. *et al.* How can soil monitoring networks be used to improve predictions of organic carbon pool dynamics and CO₂ fluxes in agricultural soils? *Plant Soil* **338**, 247–259 (2011).
81. Ogle, S. M., Breidt, F. J., Easter, M., Williams, S. & Paustian, K. An empirically based approach for estimating uncertainty associated with modeling carbon sequestration in soils. *Ecol. Modell.* **205**, 453–463 (2007).
82. Peters, W. *et al.* An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proc. Natl Acad. Sci. USA* **104**, 18925–18930 (2007).
83. Schuh, A. E. *et al.* Evaluating atmospheric CO₂ inversions at multiple scales over highly-inventoried agricultural landscape. *Glob. Change Biol.* **19**, 1424–1439 (2013).
84. Miles, N. L. *et al.* Large amplitude spatial and temporal gradients in atmospheric boundary layer CO₂ mole fractions detected with a tower-based network in the U.S. upper Midwest. *J. Geophys. Res.* **117**, G01019 (2012).
85. Lauvaux, T. *et al.* Constraining the CO₂ budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system. *Atmos. Chem. Phys.* **12**, 337–354 (2012).
86. Cooley, D., Breidt, F. J., Ogle, S. M., Schuh, A. E. & Lauvaux, T. A constrained least-squares approach to combine bottom-up and top-down CO₂ flux estimates. *Environ. Ecol. Stat.* **20**, 129–146 (2013).
- One of the first studies using independent ground- and atmosphere-based methods to validate GHG emissions estimates for a large agriculturally intensive region.**

87. Kadygrov, N. *et al.* Role of simulated GOSAT total column CO₂ observations in surface CO₂ flux uncertainty reduction. *J. Geophys. Res.* **114**, D21208 (2009).
88. Basu, S. *et al.* Global CO₂ fluxes estimated from GOSAT retrievals of total column CO₂. *Atmos. Chem. Phys.* **13**, 8695–8717 (2013).
89. Pacala, S. & Socolow, R. Stabilization wedges: solving the climate problem for the next 50 years with current technologies. *Science* **305**, 968–972 (2004).
90. Alexander, P., Paustian, K., Smith, P. & Moran, D. The economics of soil C sequestration and agricultural emissions abatement. *Soil* **1**, 331–339 (2015).
91. NitroEurope <http://www.nitroeuropa.eu> (accessed May 2015).
92. GRACEnet http://www.ars.usda.gov/research/programs/programs.htm?np_code=212&docid=21223 (USDA, accessed May 2015).
93. Dhawale, N. M. *et al.* Proximal soil sensing of soil texture and organic matter with a prototype portable mid-infrared spectrometer. *Eur. J. Soil Sci.* **66**, 661–669 (2015).
94. Yeluripati, J. *et al.* Global Research Alliance Modelling Platform (GRAMP): an open web platform for modelling greenhouse gas emissions from agro-ecosystems. *Comp. Elec. Agric.* **111**, 112–120 (2015).
95. Sanchez, P. Digital soil map of the world. *Science* **325**, 680–681 (2009).
96. The Web Soil Survey <http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx> (accessed May 2015).
97. Awan, M. I., van Oort, P. A. J., Ahmad, R., Bastiaans, L. & Meinke, H. Farmers' views on the future prospects of aerobic rice culture in Pakistan. *Land Use Policy* **42**, 517–526 (2015).
98. Smith, P. *et al.* Policy and technological constraints to implementation of greenhouse gas mitigation options in agriculture. *Agric. Ecosyst. Environ.* **118**, 6–28 (2007).
99. Louwagie, G. S.H., Sammeth, F. & Ratering, T. The potential of European Union policies to address soil degradation in agriculture. *Land Degrad. Dev.* **22**, 5–17 (2011).
100. Decision No 529/2013/EU of the European Parliament and of the Council of 21 May 2013 on accounting rules on greenhouse gas emissions and removals resulting from activities relating to land use, land-use change and forestry and on information concerning actions relating to those activities. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_2013_165_R_0080_01 (EUR-Lex, Access to European Union Law, accessed May 2015).
101. *Field to Market Alliance* <https://www.fieldtomarket.org/> (accessed May 2015).
102. *Unilever Sustainable Agriculture Code* https://www.unilever.com/Images/sac-2015_tcm244-427050_en.pdf (accessed Feb 2016).
103. Horowitz, J. K. & Just, R. E. Economics of additionality for environmental services from agriculture. *J. Environ. Econ. Manage.* **66**, 105–122 (2013).

Acknowledgements We thank A. Swan for figure design. We thank the following organisations for support: USDA/NIFA (grant number 2011-67003-30205 to K.P. and S.O.); USDA/NRCS (grant number CESU-68-7482-15-507 to K.P.); the NSF (grant number DEB 1027253 to G.P.R.); the US DOE (grant number DE-FC02-07ER64494 to G.P.R.); NERC (grant number NE/M016900/1 to P.S.); and the Belmont Forum/FACCE-JPI (grant number NE/M021327/1 to P.S.).

Author Contributions K.P. led the development of the manuscript and the integration of content. All authors contributed equally to drafting sections of the manuscript and making revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.P. (keith.paustian@colostate.edu).

Thalamic reticular impairment underlies attention deficit in *Ptchd1*^{Y/-} mice

Michael F. Wells^{1,2*}, Ralf D. Wimmer^{3,4*}, L. Ian Schmitt^{3,4}, Guoping Feng^{2,5} & Michael M. Halassa^{3,4,6,7}

Developmental disabilities, including attention-deficit hyperactivity disorder (ADHD), intellectual disability (ID), and autism spectrum disorders (ASD), affect one in six children in the USA. Recently, gene mutations in patched domain containing 1 (*PTCHD1*) have been found in ~1% of patients with ID and ASD. Individuals with *PTCHD1* deletion show symptoms of ADHD, sleep disruption, hypotonia, aggression, ASD, and ID. Although *PTCHD1* is probably critical for normal development, the connection between its deletion and the ensuing behavioural defects is poorly understood. Here we report that during early post-natal development, mouse *Ptchd1* is selectively expressed in the thalamic reticular nucleus (TRN), a group of GABAergic neurons that regulate thalamocortical transmission, sleep rhythms, and attention. *Ptchd1* deletion attenuates TRN activity through mechanisms involving small conductance calcium-dependent potassium currents (SK). TRN-restricted deletion of *Ptchd1* leads to attention deficits and hyperactivity, both of which are rescued by pharmacological augmentation of SK channel activity. Global *Ptchd1* deletion recapitulates learning impairment, hyper-aggression, and motor defects, all of which are insensitive to SK pharmacological targeting and not found in the TRN-restricted deletion mouse. This study maps clinically relevant behavioural phenotypes onto TRN dysfunction in a human disease model, while also identifying molecular and circuit targets for intervention.

Recent genetic studies have revealed substantial overlap of risk genes across seemingly distinct neurodevelopmental and psychiatric disorders including ASD, ADHD, schizophrenia, and ID^{1–4}. Such shared genetic architectures could potentially explain the overlap of behavioural abnormalities across these diagnostic categories, but because of the difficulty in mapping circuitry mechanisms of behaviour, understanding how diverse genetic lesions converge onto behaviour-relevant circuit dysfunction has been limited.

Here we focused on *PTCHD1*, a gene that is mutated in about 1% of all patients with ASD and ID^{5–10}. Comprehensive clinical analysis of *PTCHD1* deletion patients identified a variable non-syndromic neurodevelopmental disorder with symptoms ranging from attention deficit, hyperactivity, sleep abnormality, hypotonia, and learning disability¹¹. We found that *Ptchd1* was selectively expressed in the TRN of mice in early development and continued to be enriched in this structure throughout adult life. The TRN is critical for thalamocortical transmission^{12–15}, generation of sleep rhythms^{12,16–18}, sensorimotor processing^{19,20}, and attention^{13,21}, and its perturbation could result in deficits in these domains. By generating a conditional *Ptchd1*-knockout mouse, we mapped ADHD-like behaviours onto TRN circuit dysfunction via two independent methods. First, by deleting *Ptchd1* selectively from the TRN, we replicated the attention deficit and hyperactivity behaviours, but not other disease-related phenotypes found in the full knockout. Second, pharmacological rescue of TRN biophysical dysfunction selectively rescued these ADHD-related behaviours in the *Ptchd1* knockout. These findings constitute the first evidence for a ‘leaky thalamus’ in a neurodevelopmental disorder, where irrelevant inputs that are normally suppressed become highly distracting. Most importantly, we identified the TRN and its SK channels as circuit and molecular targets for intervention.

Altered TRN neuronal biophysics

The X-linked *Ptchd1* gene is predicted to encode a twelve-pass transmembrane protein with a sterol-sensing domain⁵, prompting its classification as a member of the Patched family and speculation that it may function as a Sonic hedgehog receptor^{22–24}. Prenatally, *Ptchd1* expression is found in the developing cerebellum and diencephalon (http://www.ncbi.nlm.nih.gov/nucore/NM_001093750). We found *Ptchd1* mRNA to be confined to the TRN at birth, and by post-natal day 15 (P15) onwards, to be expressed in the striatum, cortex, and cerebellum (Fig. 1a, b, Extended Data Fig. 1, Supplementary Table 1). Interestingly, Patched family members *Ptchd2* and *Ptchd3* show no TRN expression (<http://www.brain-map.org/>, 77620810 and 71891731, respectively). Thus, *Ptchd1* may play a unique role in the TRN.

To understand how *Ptchd1* deletion might contribute to neurodevelopmental disorders, we generated a conditional allele of *Ptchd1* by targeting exon 2 (Extended Data Fig. 2a). This exon encodes 3 out of the 12 transmembrane domains, including a substantial portion of the sterol-sensing domain. Loss of this exon is predicted to generate a prematurely terminated non-functional protein (Extended Data Fig. 2b). *In situ* hybridization, genotype PCR, and cDNA transcript analyses confirmed successful excision of exon 2 (Extended Data Fig. 2c–e; for source data, see Supplementary Fig. 1). Because *Ptchd1* is X-linked and individuals with *PTCHD1* deletion are almost exclusively males, we used hemizygous male mice (*Ptchd1*^{Y/-}; referred to as *Ptchd1*-knockout) for this study. *Ptchd1*-knockout mice were viable with normal body weight allowing for direct genetic modelling of this human condition.

Given the enriched expression of *Ptchd1* in the TRN, we focused our initial investigation on this structure in *Ptchd1*-knockout mice. The TRN is a group of GABAergic neurons that provide the major source of inhibition to thalamic relay nuclei, and are thought to regulate cortical

¹Department of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA. ²McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Neuroscience Institute, New York University Langone Medical Center, New York, New York 10016, USA.

⁴Department of Neuroscience and Physiology, New York University Langone Medical Center, New York, New York 10016, USA. ⁵Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁶Department of Psychiatry, New York University Langone Medical Center, New York, New York 10016, USA. ⁷Center for Neural Science, New York University, New York, New York 10003, USA.

*These authors contributed equally to this work.

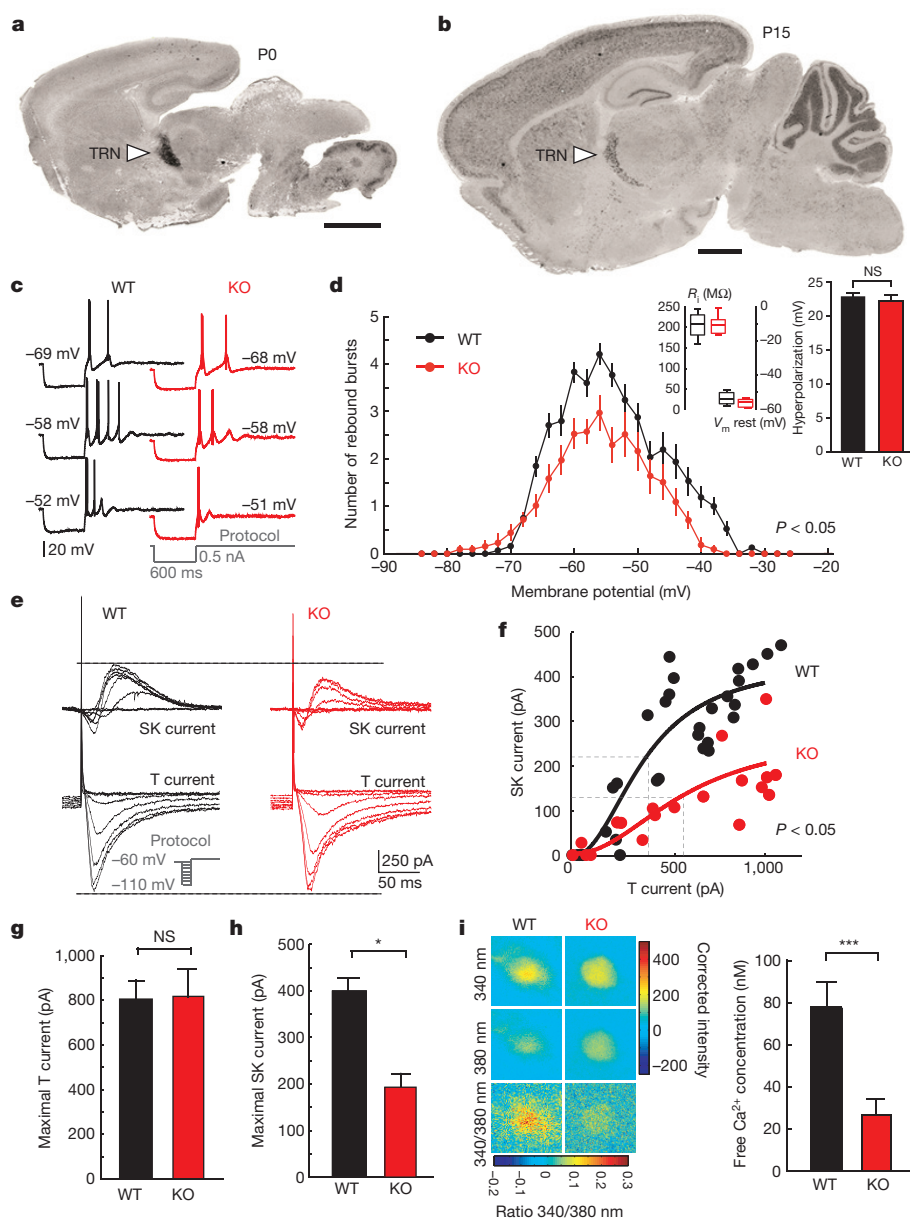


Figure 1 | Impaired repetitive bursting and SK2 currents in TRN neurons from knockout mice.

a, b, *Ptchd1* expression ($n = 3$ wild-type (WT) mice). Scale bar, 1 mm. **c**, Representative TRN burst traces ($n = 8$ WT, 9 KO cells). **d**, Reduced burst firing in knockout (KO) TRN neurons ($n = 8$ WT, 9 KO cells). V_m rest, resting membrane potential; R_i , input resistance. **e**, Representative T and SK2 current traces. **f–h**, Normal T (**g**) and reduced SK2 currents (**h**) in knockout cells ($n = 8$ WT, 9 KO). **i**, Diminished free $[Ca^{2+}]_i$ in knockout cells. Representative heat maps show background-corrected intensity ($n = 37$ WT, 36 KO cells). Wilcoxon rank-sum (**d**, **f–h**) and two-tailed t -tests (**i**). Error bars, mean \pm s.e.m. * $P < 0.05$; *** $P < 0.001$; NS, not significant.

rhythms, sleep, and attention^{21,25,26}. To begin investigating possible physiological changes in the knockout mice, we exploited a well-known characteristic of TRN neurons. Depending on their resting membrane potential, these neurons fire in two distinct modes upon receiving synaptic input²⁷. At depolarized membrane potential, they fire tonic Na^+ spikes. When hyperpolarized, they generate repetitive 'low-threshold' Ca^{2+} transients crowned by high-frequency Na^+ spikes known as bursts^{26,28–30}. Whole-cell patch-clamp recordings of TRN neurons revealed a significant decrease in repetitive bursting in knockout mice compared to wild-type controls (Fig. 1c, d). These changes were not the result of altered knockout TRN neuron resting membrane potential, input resistance, or escape from hyperpolarization (Fig. 1d, inset).

Because repetitive bursting is known to depend on interactions between T-type Ca^{2+} and small conductance calcium-activated K^+ channels²⁶, we asked which of these two conductances were primarily impaired in the knockout mice. Under voltage clamp, we found that T currents were intact but SK currents were reduced by 50% in the knockout (Fig. 1e–h). As SK channels are sensitive to resting state intracellular Ca^{2+} ($[Ca^{2+}]_i$)^{26,31}, we measured this concentration using the ratiometric Ca^{2+} indicator Fura-2AM in TRN neurons from acute brain slices. We found a twofold reduction of $[Ca^{2+}]_i$ in knockout

TRN neurons (Fig. 1i), suggesting that altered Ca^{2+} homeostasis may underlie SK channel deficits.

Reduced TRN-generated sleep spindles

Previous studies have suggested that TRN bursting plays a role in the generation of sleep spindles^{19,28}, predicting that the diminished bursting found in *Ptchd1*-knockout mice would lead to reduced sleep spindles. Using independently adjustable multi-electrode arrays to directly target TRN neurons for electrophysiological recordings in freely behaving animals and surface electroencephalography (Fig. 2a, Extended Data Fig. 3a)³², we discovered that TRN neurons from knockout mice exhibited reduced burst firing in sleep (Fig. 2b) and knockout mice showed an overall reduction in sleep spindle count (Fig. 2c, Extended Data Fig. 3b, c). Further, the degree of TRN neuronal engagement in spindle events was substantially diminished in the knockout (Extended Data Fig. 3d–f), supporting the link between the cellular and network phenotypes in this disorder and perhaps other human neurodevelopmental disorders^{33,34}. Importantly, consistent with the notion that sleep spindles are a marker for sleep stability^{35,36}, we found *Ptchd1*-knockout mice to display highly fragmented sleep (Fig. 2d–f, Supplementary Table 2).

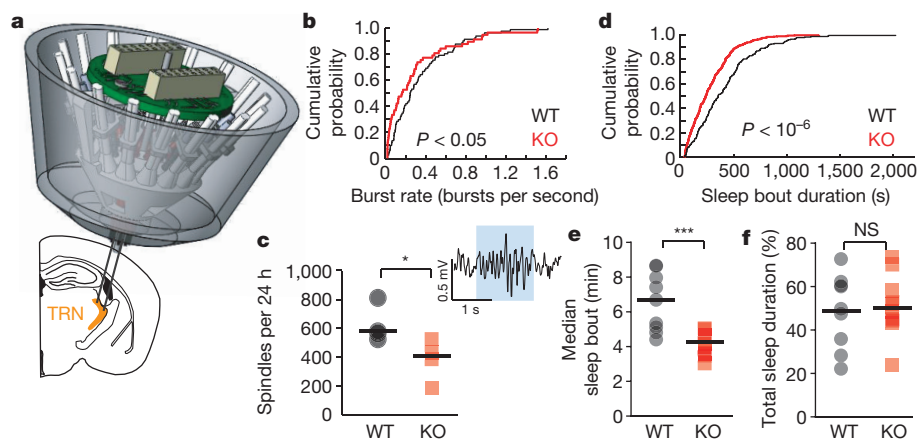


Figure 2 | Decreased spindles and sleep fragmentation in knockout mice

a, Multi-electrode implant targeting TRN. **b**, Decreased TRN burst discharge in knockouts ($n = 89$ WT, 80 KO cells from 4 WT, 3 KO mice). **c**, Reduced spindles in knockout mice. **d–f**, Knockout mice display shorter sleep bouts with normal total sleep duration ($n = 9$ WT, 10 KO). Kolomgorov–Smirnov (**b**, **d**) and Wilcoxon rank-sum tests (**c**, **e–f**). Horizontal, median (**c**, **e–f**). * $P < 0.05$; *** $P < 0.001$.

Impaired sensory-evoked thalamic inhibition

In addition to reduced rebound bursting, insufficient K^+ -mediated hyperpolarization is expected to more generally alter TRN neuronal excitability. Most importantly, it could lead to neurons not being hyperpolarized enough for T-type Ca^{2+} channels to de-inactivate and boost excitability²⁴. Therefore, despite finding T-type currents to be intact under controlled voltage-clamp conditions (at -70 mV), insufficient hyperpolarization caused by reduced SK currents would render T-type channels less recruitable under physiological conditions¹⁷, leading to reduced TRN neuronal activity and diminished overall thalamic inhibition. To test this prediction at the population level and in the intact brain, we used chloride photometry, a tool that we recently developed as a proxy for population-level GABAergic inhibition¹³. This technique utilizes fluorescence resonance energy transfer (FRET)-based measurements of the chloride-sensor SuperClomeleon³⁷, a reporter composed of a cyan fluorescent protein (CFP) FRET-donor and a chloride-quenchable yellow fluorescent protein (YFP) FRET acceptor (Fig. 3a). By introducing SuperClomeleon into visual thalamic neurons (lateral geniculate nucleus, LGN; Fig. 3b), we observed visual-evoked chloride transients (Fig. 3c, d), replicating our recent findings¹³. Visual-evoked inhibitory transients were observed in both wild-type and *Ptchd1*-knockout LGN, but a quantitative comparison revealed a 25% reduction of these transients in

the knockout (Fig. 3e–g). Impaired inhibition was also observed in response to trains of stimuli (Fig. 3h, j). Interestingly, a small inhibitory augmentation as a result of repeated stimulation was observed in wild-type mice and this augmentation was also significantly reduced in the knockouts (Fig. 3h, i). Altogether, these findings provide direct evidence for reduced thalamic inhibition in *Ptchd1*-knockout mice and are consistent with impaired TRN output with ensuing deficits in sensory-related thalamic inhibition.

ADHD-like behaviours in *Ptchd1*-knockout mice

We have previously observed that thalamic inhibition is used to suppress unwanted sensory inputs during attention¹². The observed reduction in thalamic inhibition (Fig. 3) would predict that unwanted sensory inputs may become particularly distracting for *Ptchd1*-knockout mice. To test this hypothesis, we trained mice on a visual detection task shown to require attentional engagement¹² (Fig. 4a). Mice initiated each trial by continuously breaking an infrared barrier for 0.5–0.7 s, ensuring proper head position when a visual stimulus was presented either to the right or left of the animal. Correct indication of visual stimulus location by nose-poking resulted in reward delivery. Under such conditions, *Ptchd1*-knockout mice performance was comparable to that of a wild type (Fig. 4b). However, in the presence of a visual distractor during anticipation, knockout mice showed impaired

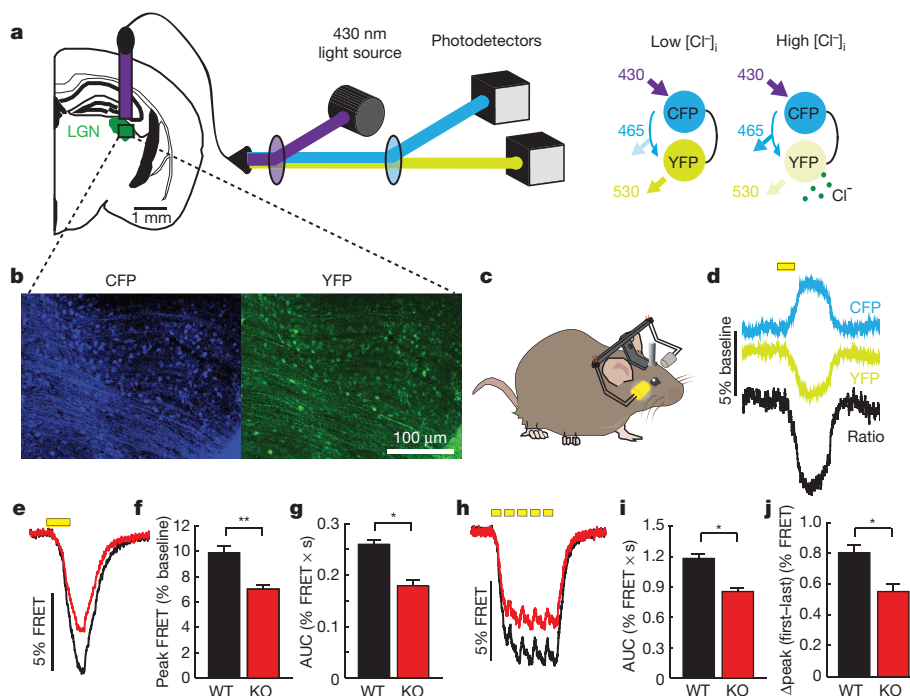


Figure 3 | Reduced sensory-evoked thalamic inhibition in knockout mice.

a, Schematic of CFP to YFP FRET. **b**, Confocal images of SuperClomeleon expression in LGN. **c**, Stimuli delivered to the eye contralateral to implanted LGN. **d**, Example traces of visual-evoked CFP and YFP fluorescence changes. **e–g**, Reduced LGN inhibition in knockouts, reflected in peak FRET response (**f**) and smaller area under the curve (AUC; **g**). **h–j**, Knockout mice also show decreased facilitation of FRET response. $n = 6$ WT, 6 KO mice. Wilcoxon rank-sum tests (**e–j**). Error bars, mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$.

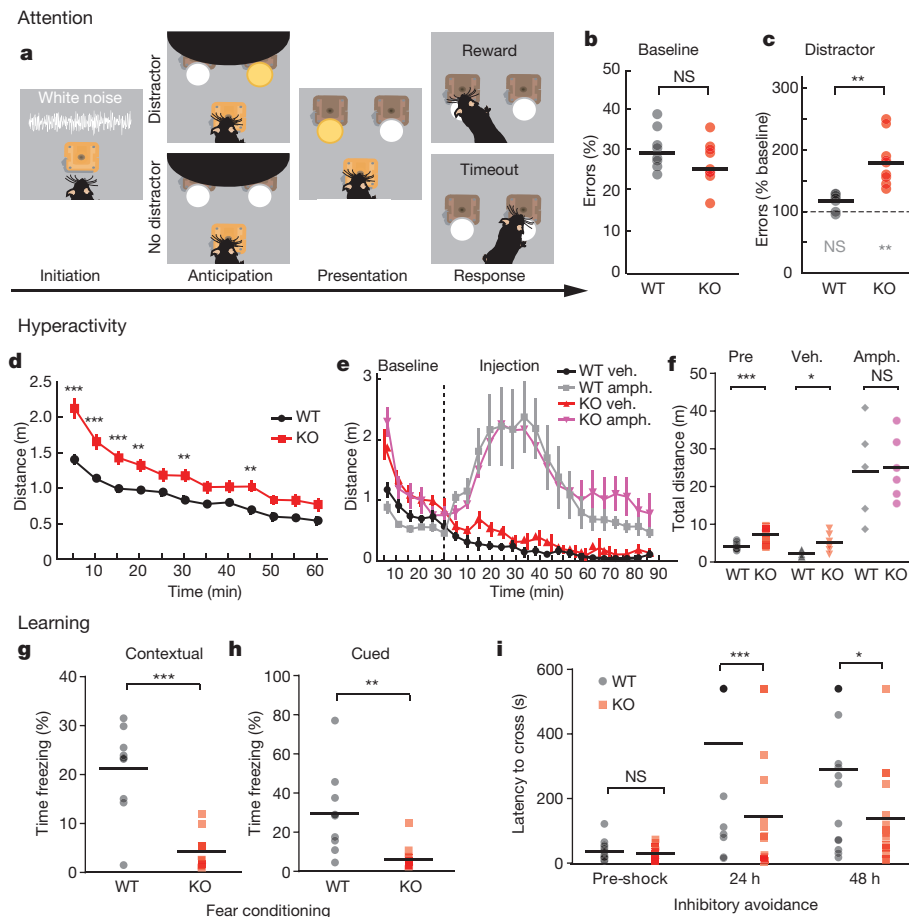


Figure 4 | Knockout mice show attention, locomotor, and learning impairment. **a**, Visual detection task design ($n = 8$ WT, 9 KO). **b**, Knockouts showed comparable baseline performance. **c**, Knockouts displayed decreased accuracy in the presence of distractors. Dashed line, baseline performance (without distractors). **d**, Knockouts show increased locomotion in open field ($n = 30$ WT, 31 KO). **e**, **f**, Knockouts show normal responses to amphetamine ($n = 5$ vehicle-treated WT (WT veh.), 6 amphetamine-treated WT (WT amph.), 5 KO veh., 6 KO amph.). **g**, **h**, Knockouts exhibit decreased fear-induced learning behaviours in contextual (**g**) and cued (**h**) fear conditioning tests ($n = 10$ WT, 11 KO). **i**, Diminished knockout latency to cross in inhibitory avoidance task ($n = 24$ WT, 23 KO). Wilcoxon rank-sum (**b**, **c**), two-tailed t -test (**g**, **h**) and two-way repeated measures ANOVA with Bonferroni post-hoc tests (**d**–**f**, **i**). Horizontal bars, median (**b**, **c**), mean (**f**–**i**); error bars, mean \pm s.e.m. (**d**, **e**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (grey stars in **c** indicate statistical difference compared to baseline performance).

performance (Fig. 4c). The specific distractibility phenotype, rather than a more general failure of attentional engagement, revealed by this novel behavioural task is consistent with the prediction of impaired thalamic inhibition required for distractor suppression. It is also consistent with clinical findings in patients with related neurodevelopmental disorders³⁸. This impairment was not the result of general sensorimotor dysfunction, given the intact performance on standard sensorimotor testing (Extended Data Fig. 4a–c).

Distractibility is often accompanied by hyperactivity in several human neurodevelopmental disorders such as ADHD^{39,40}. ADHD symptoms are frequently observed in patients with *PTCHD1* mutations¹¹. Interestingly, *Ptchd1*-knockout mice showed a hyperactivity phenotype in the open field (Fig. 4d). Classical ADHD-related hyperactivity is predicted to be treated effectively with amphetamines, as has been previously described in other mouse models of neurodevelopmental disease⁴¹. Surprisingly, *Ptchd1*-knockout hyperactivity was insensitive to amphetamine treatment (Fig. 4e, f), suggesting a unique pathophysiological origin that may be related to the approximately 30% of ADHD patients who do not respond to amphetamines⁴².

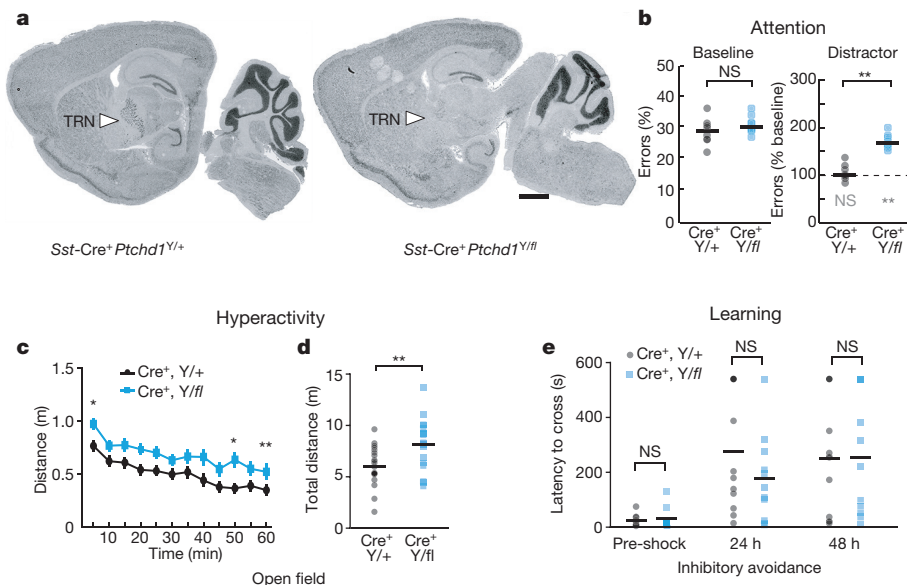
In addition to attention deficits and hyperactivity, *Ptchd1*-knockout mice showed a variety of behavioural abnormalities. Although knockout mice showed intact performance on tasks requiring simple spatial learning (Extended Data Fig. 5a–d), they exhibited significant deficits on tasks necessitating more complex associations that are believed to require integration across multiple brain structures⁴³. Knockout mice showed fear-induced freezing deficits in contextual (Fig. 4g) and cued (Fig. 4h) fear-conditioning tests. Impaired learning was corroborated by profound deficits on the inhibitory avoidance task (Fig. 4i). Knockout mice also exhibited motor defects such as gait abnormalities and hypotonia, as well as hyper-aggression (Extended Data Fig. 5e–h). All of these behaviours were independent of genetic background (Extended Data Fig. 6). These behaviours are consistent with clinical

findings of multi-system abnormalities in *PTCHD1* deletion patients, suggesting the major function of *PTCHD1* is evolutionarily conserved. Interestingly, although several *PTCHD1* deletion patients have been diagnosed with ASD, knockout mice did not exhibit differences in repetitive grooming (Extended Data Fig. 7a) or social interaction (Extended Data Fig. 7b, c). This may reflect evolutionary divergence either in *PTCHD1* function or in behavioural circuits related to *PTCHD1* deficiency⁴⁴.

TRN defects underlie ADHD-like behaviours

To determine which behavioural abnormalities are caused by TRN dysfunction, we sought strategies to largely limit *Ptchd1* deletion to the TRN. To choose the appropriate Cre-driver line to breed with floxed *Ptchd1* (*Ptchd1*^{+/f}) mice, we looked for specific TRN overlap between candidate markers and *PTCHD1* expression. Immunohistochemical co-labelling experiments of *Ptchd1*-YFP^{+/+}, a novel knock-in mouse with yellow fluorescent protein (YFP) in place of *Ptchd1* exon 1 (Extended Data Fig. 8a), revealed significant and unique overlap between YFP and the inhibitory neuronal marker GAD67 in the TRN (Extended Data Fig. 8b, Supplementary Table 3).

Both parvalbumin and somatostatin are inhibitory neural markers that are widely expressed across TRN neurons^{45,46} and showed overlap with YFP limited to this region (Extended Data Fig. 8c), suggesting that mating Cre-drivers of either parvalbumin or somatostatin to *Ptchd1*^{+/f} mice would result in a *Ptchd1* deletion that is primarily confined to the TRN. Given the early post-natal expression of *PTCHD1* and somatostatin in the TRN (Extended Data Fig. 9), we reasoned that knock-in mice expressing Cre recombinase in somatostatin neurons without disrupting endogenous *Sst* expression (*Sst*-Cre) would be a more suitable choice for such experiments⁴⁷. We crossed *Ptchd1*^{+/f} female mice to *Sst*-Cre mice and generated male mice lacking *Ptchd1* in the TRN (*Sst*-Cre⁺ *Ptchd1*^{Y/f}), as confirmed by *in situ* hybridization (Fig. 5a).



Sst-Cre⁺ Ptchd1^{Y/f} mice displayed attention deficits (Fig. 5b) and recapitulated the hyperactivity phenotype observed in the germline knockout (Fig. 5c, d), suggesting that these behaviours are explained by TRN dysfunction. *Sst-Cre⁺ Ptchd1^{Y/f}* mice showed intact learning (Fig. 5e) and did not show hypotonia or hyper-aggression (Extended Data Fig. 10a, b), reaffirming the circuit specificity of this genetic model. Interestingly, *Sst-Cre⁺ Ptchd1^{Y/f}* mice exhibited fragmented sleep (Extended Data Fig. 10c–e), confirming the notion that sleep abnormalities observed in the germline knockout are of TRN origin, and that sleep and attention deficits can arise from common circuit dysfunction⁴⁸.

Finally, we asked whether pharmacological boosting of SK channels could rescue ADHD-like knockout behaviours. Acute injection of the SK positive allosteric modulator 1-ethyl-benzimidazolinone (1-EBIO) significantly mitigated impaired sensory-evoked thalamic inhibition in germline knockout mice with no impact on inhibitory transients in the wild types (Fig. 6a). Consistent with these physiological effects, 1-EBIO did not affect attentional task performance in the wild-type mice (Fig. 6b), but substantially mitigated distractibility in the knockout mice (Fig. 6c). In addition, 1-EBIO injection rescued the hyperactivity phenotype in the knockouts (Fig. 6d, e). The specificity of this pharmacological approach was supported by its lack of effect on other abnormalities including hypotonia (Extended Data Fig. 10f),

hyper-aggression (Extended Data Fig. 10g), and learning deficits (Fig. 6f). Together, these results further support SK channel dysfunction as a cellular mechanism for these behavioural abnormalities. Future studies examining the therapeutic benefit of SK targeting for sleep fragmentation and instability and its potential relevance to inattention would be important.

Discussion

To our knowledge, this study is the first to show that a TRN circuit deficit is central to a specific set of behavioural impairments in a human neurodevelopmental disease model. Using conditional knockout of *Ptchd1*, a gene with expression restricted to the TRN during early post-natal development, we mapped behavioural phenotypes onto their circuit substrates. We additionally discovered that modulation of SK channel function could be explored as a potential novel treatment strategy for *PTCHD1* deletion patients with attention deficits and hyperactivity. Hyperactivity of TRN origin may be the motor equivalent of sensory distractibility, which could involve dysfunctional motor TRN-thalamic circuits. Future experiments exploring inhibitory control of motor thalamus will formally test this conjecture.

Although basic studies have shown the TRN to be central for attention^{21,49} and sleep spindles¹⁶, our study directly shows how disease-relevant impaired TRN output can result in attention deficits,

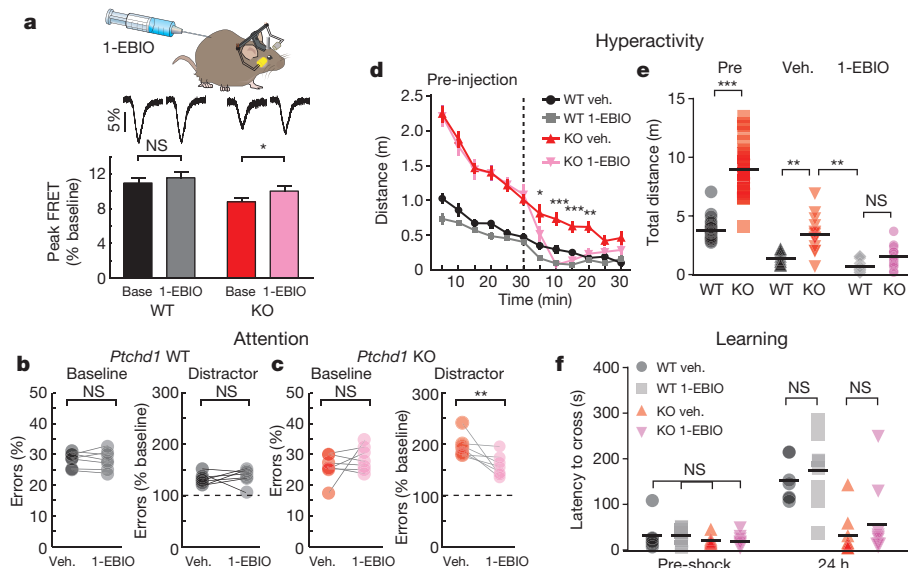


Figure 6 | SK conductance augmentation corrects ADHD-like symptoms in knockout mice. **a**, 1-EBIO

corrects inhibitory transients in knockouts ($n = 7$ WT veh., 7 1-EBIO-treated WT (WT 1-EBIO), 7 KO veh., 7 KO 1-EBIO). **b**, **c**, 1-EBIO improves attention performance in knockout mice ($n = 7$ WT veh., 7 WT 1-EBIO, 7 KO veh., 7 KO 1-EBIO). Dashed lines, baseline performance; grey bars connect data points from the same animal. **d**, **e**, 1-EBIO rescues hyperactivity in knockouts ($n = 13$ WT veh., 13 WT 1-EBIO, 13 KO veh., 15 KO 1-EBIO). **f**, Treatment with 1-EBIO did not affect inhibitory avoidance ($n = 7$ WT veh., 8 WT 1-EBIO, 9 KO veh., 10 KO 1-EBIO). Wilcoxon rank-sum (**a**–**c**) and two-way reduced measures ANOVA with Bonferroni post-hoc tests (**d**–**f**). Error bars, mean \pm s.e.m. (**a**, **d**); horizontal bars, mean (**e**, **f**). **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

hyperactivity, and sleep disruption. This direct demonstration was made possible by developing SuperClomeleon photometry, a technique that can now be widely applied to disease models as a screen for impaired thalamic inhibition. It is possible that a 'leaky thalamus' caused by impaired TRN function underlies attention deficits, hyperactivity, and sleep disruption across various neurodevelopmental disorders, and we expect the set of genetic, physiologic, and behavioural approaches we introduce here to facilitate such discoveries.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 October 2015; accepted 17 February 2016.

Published online 23 March 2016.

- Coe, B. P., Girirajan, S. & Eichler, E. E. The genetic variability and commonality of neurodevelopmental disease. *Am. J. Med. Genet. C. Semin. Med. Genet.* **160C**, 118–129 (2012).
- Coe, B. P., Girirajan, S. & Eichler, E. E. A genetic model for neurodevelopmental disease. *Curr. Opin. Neurobiol.* **22**, 829–836 (2012).
- Zhou, Y. *et al.* Mice with *Shank3* mutations associated with ASD and schizophrenia display both shared and distinct defects. *Neuron* **89**, 147–162 (2016).
- Cristino, A. S. *et al.* Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol. Psychiatry* **19**, 294–301 (2014).
- Noor, A. *et al.* Disruption at the *PTCHD1* locus on Xp22.11 in Autism spectrum disorder and intellectual disability. *Sci. Transl. Med.* **2**, 49ra68 (2010).
- Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
- Whibley, A. C. *et al.* Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability. *Am. J. Hum. Genet.* **87**, 173–188 (2010).
- Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
- Filges, I. *et al.* Deletion in Xp22.11: *PTCHD1* is a candidate gene for X-linked intellectual disability with or without autism. *Clin. Genet.* **79**, 79–85 (2011).
- Torricco, B. *et al.* Contribution of common and rare variants of the *PTCHD1* gene to autism spectrum disorders and intellectual disability. *Eur. J. Hum. Genet.* **23**, 1694–1701 (2015).
- Chaudhry, A. *et al.* Phenotypic spectrum associated with *PTCHD1* deletions and truncating mutations includes intellectual disability and autism spectrum disorder. *Clin. Genet.* **88**, 224–233 (2015).
- Halassa, M. M. *et al.* State-dependent architecture of thalamic reticular subnetworks. *Cell* **158**, 808–821 (2014).
- Wimmer, R. D. *et al.* Thalamic control of sensory selection in divided attention. *Nature* **526**, 705–709 (2015).
- Pinault, D. The thalamic reticular nucleus: structure, function and concept. *Brain Res. Brain Res. Rev.* **46**, 1–31 (2004).
- Guillery, R. W., Feig, S. L. & Lozsadi, D. A. Paying attention to the thalamic reticular nucleus. *Trends Neurosci.* **21**, 28–32 (1998).
- Halassa, M. M. *et al.* Selective optical drive of thalamic reticular nucleus generates thalamic bursts and cortical spindles. *Nature Neurosci.* **14**, 1118–1120 (2011).
- Barthó, P. *et al.* Ongoing network state controls the length of sleep spindles via inhibitory activity. *Neuron* **82**, 1367–1379 (2014).
- von Krosigk, M., Bal, T. & McCormick, D. A. Cellular mechanisms of a synchronized oscillation in the thalamus. *Science* **261**, 361–364 (1993).
- Marlinski, V., Sirota, M. G. & Beloozerova, I. N. Differential gating of thalamocortical signals by reticular nucleus of thalamus during locomotion. *J. Neurosci.* **32**, 15823–15836 (2012).
- Erlj, D. *et al.* Dopamine D4 receptor stimulation in GABAergic projections of the globus pallidus to the reticular thalamic nucleus and the substantia nigra reticulata of the rat decreases locomotor activity. *Neuropharmacology* **62**, 1111–1118 (2012).
- McAlonan, K., Cavanaugh, J. & Wurtz, R. H. Guarding the gateway to cortex with attention in visual thalamus. *Nature* **456**, 391–394 (2008).
- Zhong, Y. *et al.* Comprehensive analysis of patched domain-containing genes reveals a unique evolutionary pattern. *Genet. Mol. Res.* **13**, 7318–7331 (2014).
- Goodrich, L. V. *et al.* Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109–1113 (1997).
- Rohatgi, R., Milenkovic, L. & Scott, M. P. Patched1 regulates hedgehog signaling at the primary cilium. *Science* **317**, 372–376 (2007).
- Jasper, H. Diffuse projection systems: the integrative action of the thalamic reticular system. *Electroencephalogr. Clin. Neurophysiol.* **1**, 405–419 (1949).
- Cueni, L. *et al.* T-type Ca²⁺ channels, SK2 channels and SERCAs gate sleep-related oscillations in thalamic dendrites. *Nature Neurosci.* **11**, 683–692 (2008).
- Jahnsen, H. & Llinas, R. Voltage-dependent burst-to-tonic switching of thalamic cell activity: an in vitro study. *Arch. Ital. Biol.* **122**, 73–82 (1984).
- Astori, S. *et al.* The Ca_v3.3 calcium channel is the major sleep spindle pacemaker in thalamus. *Proc. Natl Acad. Sci. USA* **108**, 13823–13828 (2011).
- Huguenard, J. R. & Prince, D. A. A novel T-type current underlies prolonged Ca²⁺-dependent burst firing in GABAergic neurons of rat thalamic reticular nucleus. *J. Neurosci.* **12**, 3804–3817 (1992).
- Ying, S. W. & Goldstein, P. A. Propofol-block of SK channels in reticular thalamic neurons enhances GABAergic inhibition in relay neurons. *J. Neurophysiol.* **93**, 1935–1948 (2005).
- Coulon, P. *et al.* Burst discharges in neurons of the thalamic reticular nucleus are shaped by calcium-induced calcium release. *Cell Calcium* **46**, 333–346 (2009).
- Brunetti, P. M. *et al.* Design and fabrication of ultralight weight, adjustable multi-electrode probes for electrophysiological recordings in mice. *J. Vis. Exp.* **91**, e51675 (2014).
- Ferrarelli, F. *et al.* Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* **164**, 483–492 (2007).
- Limoges, E. *et al.* Atypical sleep architecture and the autism phenotype. *Brain* **128**, 1049–1061 (2005).
- Dang-Vu, T. T. *et al.* Spontaneous brain rhythms predict sleep stability in the face of noise. *Curr. Biol.* **20**, R626–R627 (2010).
- Wimmer, R. D. *et al.* Sustaining Sleep Spindles through Enhanced SK2-Channel Activity Consolidates Sleep and Elevates Arousal Threshold. *J. Neurosci.* **32**, 13917–13928 (2012).
- Grimley, J. S. *et al.* Visualization of synaptic inhibition with an optogenetic sensor developed by cell-free protein engineering automation. *J. Neurosci.* **33**, 16297–16309 (2013).
- Remington, A. *et al.* Selective attention and perceptual load in autism spectrum disorder. *Psychol. Sci.* **20**, 1388–1393 (2009).
- Sachs, G. S. *et al.* Comorbidity of attention deficit hyperactivity disorder with early- and late-onset bipolar disorder. *Am. J. Psychiatry* **157**, 466–468 (2000).
- Leyfer, O. T. *et al.* Comorbid psychiatric disorders in children with autism: interview development and rates of disorders. *J. Autism Dev. Disord.* **36**, 849–861 (2006).
- Won, H. *et al.* GIT1 is associated with ADHD in humans and ADHD-like behaviors in mice. *Nature Med.* **17**, 566–572 (2011).
- Spencer, T. *et al.* Efficacy of a mixed amphetamine salts compound in adults with attention-deficit/hyperactivity disorder. *Arch. Gen. Psychiatry* **58**, 775–782 (2001).
- Curzon, P., Rustay, N. R. & Browman, K. E. In *Methods of Behavior Analysis in Neuroscience* 2nd edn (ed. Buccafusco, J. J.) Ch. 2 (Boca Raton, 2009).
- Nestler, E. J. & Hyman, S. E. Animal models of neuropsychiatric disorders. *Nature Neurosci.* **13**, 1161–1169 (2010).
- Meyer, A. H. *et al.* In vivo labeling of parvalbumin-positive interneurons and analysis of electrical coupling in identified neurons. *J. Neurosci.* **22**, 7055–7064 (2002).
- Graybiel, A. M. & Elde, R. P. Somatostatin-like immunoreactivity characterizes neurons of the nucleus reticularis thalami in the cat and monkey. *J. Neurosci.* **3**, 1308–1321 (1983).
- Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
- Chen, Z. *et al.* Thalamic circuit mechanisms link sensory processing in sleep and attention. *Front. in Neural Circuits* <http://dx.doi.org/10.3389/fncir.2015.00083> (2015).
- Zikopoulos, B. & Barbas, H. Pathways for emotions and attention converge on the thalamic reticular nucleus in primates. *J. Neurosci.* **32**, 5338–5350 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Tang for insightful discussion during the initiation of the project, H. Wang, T. Dalia, E. Kwan, H. Zaniwski for technical support, and J. Vincent for insightful discussion. We thank J. Petravic and T. Emery from the Sur laboratory for assistance with Ca²⁺ imaging and A. Heynen from the Bear laboratory for technical advice on the inhibitory avoidance task. We thank all members of the Feng laboratory for their help and support. We thank M. Ball and J. Ball for their insight and inspiration throughout this project. We also thank S.F. Lin and R. Buxton for their support of this research. This work was supported by a grant from Simons Foundation Autism Research Initiative (SFARI Award ID: 307913) to G.F. and M.M.H., NIH grants to G.F. (NIH/NIMH, R01MH097104) and M.M.H. (R01MH107680), and funds from the Poitras Center for Affective Disorders Research and the Stanley Center for Psychiatric Research at the Broad Institute of MIT and Harvard to G.F. M.M.H. is additionally supported by the Brain and Behavior, Sloan, Klingenstein and Feldstein Foundations. M.F.W. is supported by an NIH Ruth L. Kirschstein National Research Service Award (FMH098641A). R.D.W. is supported by the Swiss National Science Foundation.

Author Contributions M.F.W. and G.F. conceived the genetic studies and designed associated experiments; R.D.W. and M.M.H. conceived the physiologic studies and designed associated experiments. All authors designed the behavioural studies. M.F.W. and R.D.W. collected the data. M.F.W., R.D.W. and L.I.S. analysed the data. M.F.W., R.D.W., M.M.H. and G.F. interpreted the results. M.F.W., M.M.H. and G.F. wrote the paper with input from R.D.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.F. (fengg@mit.edu).

METHODS

***Ptchd1* conditional knockout mice.** *Ptchd1* conditional knockout mice were generated by homologous recombination in R1 embryonic stem cells and implanted in C57Bl/6J blastocysts using standard procedures. The targeting vector was designed to flank exon 2 of the *Ptchd1* gene with *loxP* sites and a neomycin (Neo) selection cassette. Chimaeric mice were crossed to C57Bl/6J females (Jackson Labs). Germline transmission was determined by genotyping PCR of mouse tail DNA, using primer pFW Pt1cKO Gen 1a (5'-GGATGGTACCACCTACAATATGC-3') and pRV Pt1cKO Gen 3b (5'-AAGCCAAAGAGTTTACCCTG-3') for the wild-type allele (187 base pairs (bp)) and the floxed allele (227 bp). The F1 hybrids were crossed to C57Bl/6J β -actin Flp mice to excise the Neo cassette. The floxed mice were then backcrossed to C57Bl/6J mice for five generations. After the fifth generation, speed congenic genotyping PCRs were conducted to determine the approximate purity of the background. Only mice showing >95% C57Bl/6J background were used for subsequent matings. Backcrossed *Ptchd1*-knockout mice were then bred with C57Bl/6J β -actin Cre mice to produce germline knockouts of the floxed allele. Genotypes were determined by PCRs using the pFW Pt1cKO Gen 1a and pRV Pt1cKO Gen 4c (5'-GGACTTTGGAGTAAACCACC-3'; in the Neo cassette) primers for the knockout allele (351 bp). For all behavioural experiments in the C57Bl/6J background, *Ptchd1*^{Y/-} (knockout) and *Ptchd1*^{Y/+} (wild type) males were bred with *Ptchd1*^{+/+} females. For the C57 and 129 mixed background behavioural experiments, 129 wild-type males were bred with *Ptchd1*^{+/+} C57Bl/6J females and the F1 offspring were used for experiments. Animals were housed 3–5 by genotype per cage at a constant 23 °C in a 12 h light–dark cycle (lights on at 07:00) with *ad libitum* food and water (unless otherwise noted).

Healthy age-matched male mice between 2–5 months were used for all behavioural experiments. Sample sizes were chosen based on standards in the field as well as previous experience with phenotype comparisons. No statistical methods were used to predetermine sample size. All behaviour experiments employed a group counterbalancing strategy to ensure that both genotypes were being tested at equal frequencies throughout the testing day. For drug studies, mice were chosen at random to be included in the vehicle-treated versus drug-treated groups. Experimenters were blind to genotype, but not group (that is, group A versus group B), during data acquisition and analysis. D'Agostino and Pearson omnibus normality tests were run. When normality was violated, non-parametric tests (Wilcoxon rank-sum) were used. ANOVA was used to estimate group variance where multiple comparisons were made. All data analysed with two-tailed *t*-tests showed similar variances between groups. Data presented as scatter plots with mean denoted by bar, unless otherwise noted. All animal experiments were conducted according to the NIH Health Guide for Care and Use of Laboratory Animals and were approved by the Massachusetts Institute of Technology Institutional Animal Care and Use Committee.

***Ptchd1*-YFP mice.** *Ptchd1*-YFP mice were generated by homologous recombination in R1 embryonic stem (ES) cells and implanted in C57Bl/6J blastocysts using standard procedures. The targeting vector replaced exon 1 of the mouse *Ptchd1* gene with ATG-YFP-STOP cassette and a Neo cassette. Correctly altered ES cell colonies were PCR screened using primers targeting the YFP insert and long-arm PCR (LA Taq) methods. Chimaeric mice were crossed to C57Bl/6J females from Jackson Labs. Germline transmission was assessed through genotyping PCR of mouse tail DNA, using primers pFW Pt1-YFPki Gen 3a (5'-TTACTTCCTTTTCCCCACC-3'), pRV Pt1-YFPki Gen 3a (5'-CCCAGTTCTCTAGTAGATTCC-3'), and pFW Pt1-YFPki Gen 3b (5'-GGAGAGTGAATCTAGGTGGAG-3') for the wild-type allele (positive band, 190 bp) and the YFP positive allele (positive band, 316 bp). The F1 hybrids were backcrossed to C57Bl/6J mice for two generations. Female mice containing the YFP insertion (*Ptchd1*-YFP^{+/+}) were used for all immunohistochemistry experiments.

***In situ* hybridization.** mRNA *in situ* hybridization was performed with 20 μ m cryosections from freshly frozen P0, P15, and P35 brain tissue from male mice using a mixture of two digoxigenin (DIG)-labelled probes against mouse *Ptchd1* cDNA (GenBank Accession NM_001093750.1; ex2 base pairs 372–1006 and ex3 base pairs 1290–2027), except for the wild type and knockout comparisons, in which only the ex2 probe was used. The hybridization signal was detected using an alkaline-phosphatase-conjugated anti-DIG antibody (Roche) and developed using 5-bromo-4-chloro-indolylphosphate/nitroblue tetrazolium (Roche). Sections were imaged using an Olympus BX61 motorized fluorescent microscope.

Immunohistochemistry. Mice were perfused with ice-cold 1 \times PBS and 4% paraformaldehyde. Brains were then fixed overnight at 4 °C before vibratome sectioning. We washed 50 μ m sections with 1 \times PBS for 3 \times 5 min before a 1 h room temperature incubation in blocking solution (5% normal goat serum, 2% BSA, 0.2% Triton X-100 in 1 \times PBS). Sections were then incubated overnight at 4 °C in primary antibody solution: rabbit anti-GFP (Invitrogen A11122; 1:1000), chicken anti-GFP (Abcam 13970; 1:1000), mouse anti-GAD67 (Millipore 5406; 1:1000), rabbit anti-somatostatin (Peninsula Labs T4102; 1:1000), and mouse anti-NeuN

(Millipore MAB377; 1:1000). Following the 3 \times 20 min wash in 1 \times PBS, tissue was incubated for 4 h at room temperature in a second antibody solution: goat anti-rabbit Alexa 488 (LT A-11034; 1:1000), goat anti-rabbit Alexa 555 (LT A-21428; 1:1000), goat anti-mouse Alexa 488 (LT A-11001; 1:1000), goat anti-mouse Alexa 555 (LT A-21422; 1:1000), and goat anti-chicken Alexa 488 (LT A-11039; 1:1000). Sections were once again washed 3 \times 20 min in 1 \times PBS before mounting onto glass slides using Fluoro-Gel. Images were taken with an Olympus Fluoview 1000 confocal microscope. YFP co-labelling was quantified using ImageJ. Images were converted to 8-bit files and automatically background subtracted and thresholded. The resulting images were converted to a mask and then the processes 'fill holes' and 'watershed' were implemented. The processed images were then analysed using the 'analyze particles' feature with the squared size set to '200-Infinity' and the show feature set to 'outlines'. This end product of this procedure was an image containing enumerated cell skeletons. Skeleton images from the green and red channels were merged and overlapping skeletons were counted as a co-labelling event.

Slice electrophysiology. Freshly prepared brain slices from P21–P28 male mice were superfused with oxygenated ACSF (125 mM NaCl, 25 mM NaHCO₃, 25 mM glucose, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 1.2 mM MgCl₂, 2 mM CaCl₂, 1.7 mM ascorbic acid) and recorded at 30–34 °C. Patch pipettes (2.5–4 M Ω) contained the following intracellular solution in 140 mM KMeSO₄, 10 mM KCl, 10 mM HEPES, 0.1 mM EGTA, 4 mM Mg-ATP, 0.2 mM Na-GTP, 10 mM phosphocreatine, (285 mM mOsm, pH 7.2). For rebound burst characterization, cells were held in current clamp at potentials ranging from –85 to –25 mV through constant current injection. Rebound bursting was determined following a 600 ms, –0.5 nA current step. Events crossing two standard deviations of the baseline noise were considered bursts. To determine T and SK currents, responses to different hyperpolarizing steps (500 ms, ranging from –110 to –60 mV) were recorded in voltage clamp configuration. T currents were isolated through application of the SK channel blocker apamin (100 nM), whereas SK currents were estimated by digital subtraction of the isolated T current from control currents without apamin in slices from 4–5 mice per genotype. Cells where the input resistance changed >15% during the recording were excluded from the analysis.

***In vivo* TRN recordings.** Hyperdrives containing 12 individually adjustable microdrives loaded with 1–2 stereotrodes were built as previously described³⁷. Mice were anaesthetized with 1% isoflurane and mounted on a stereotaxic frame. A 3 \times 2.5 mm craniotomy (centre coordinate from bregma, medio-lateral: 2.5 mm, anterior-posterior: –1.3 mm) was drilled and the hyperdrive was implanted at a 15° angle relative to midline. At time of implantation, stereotrodes were lowered <500 μ m into the brain. Three stainless steel screws (one located prefrontal and two cerebellar) served as electroencephalography (EEG) electrodes and ground and anchored together with two additional fixation screws the hyperdrive to the skull. After recovery, mice were connected to a custom-made 64-channel preamplifier headstage (Neuralynx) and data was acquired using a Neuralynx Digilynx recording system. Stereotrode signals were amplified, filtered between 0.1 Hz and 9 kHz and digitized at approximately 30 kHz. Spikes were manually clustered using the MClust toolbox for MATLAB, and bursts were identified as at least two spikes with an inter-spike interval of \leq 10 ms, which were preceded by \geq 70 ms of silence. Spindle detection was performed as previously described¹². The EEG signal was filtered within the spindle frequency band (9–15 Hz) and the Hilbert transform (MATLAB function 'hilbert') was computed. The envelope of the signal (1 s smoothing) was used as a basis for spindle detection. A threshold of one standard deviation (s.d.) was applied and each threshold crossing, with parameters of >0.5 s and <3 s, were initially included. These events were subsequently visually inspected before being included in the analysis. To assess consistency in the phase locking of individual TRN neurons with spindles, we used spike phase synchrony analysis between unit firing and spindle band (9–15 Hz) filtered component of the local field potential (Extended Data Fig. 3d–f). For each TRN unit, we first constructed a spike-phase histogram of firing rates relative to alpha. Units were considered significantly phase-locked if the distribution of spike phases significantly differed from a uniform distribution on the basis of Rayleigh's test for circular uniformity (cut-off, $P < 0.05$). To quantify the degree of phase-locking, the Pearson's correlation coefficient between spike counts and phase angle was computed for each unit. Analysis was performed in MATLAB using the circular statistics toolbox.

Sleep analysis. Sleep-related immobility was used to measure sleep bout duration in mice (C57, $n = 9$ WT and 10 KO; *Sst*-Cre, $n = 10$ WT and 10 KO), a technique that has previously been shown to correlate ($r = 0.94$) with EEG recordings⁵⁰. We used BIOBEHAVE Behavioural Sequencer boxes (BIOBEHAVE) that utilize Piezo sensors and video recordings to automatically score 23 different movements/behaviours, including immobility (Sequencer label 'still'). The lighting inside the boxes was set to 5–10 lx, which mimicked home cage light intensity in the MIT animal holding facility. Mice were placed into a custom-made transparent plastic cylinder (16 cm diameter \times 16 cm height) containing home cage bedding in order to reduce the arena parameters and facilitate sleep. At 07:00, mice were placed in the

behavioural sequencers and given 1 h to acclimate. After the acclimation period, behaviour was automatically scored for 6 h (08:00–14:00). Data was manually curated to identify bouts of sleep, which were defined as at least 40 s of immobility⁵⁰ that is not interrupted by more than 2 s of upper body movements (sequencer labels 'head', 'face', 'nose', 'cheek', 'back', 'orient_look', 'orient_sniff') or by 4 or more seconds of general body movements (sequencer labels 'paw', 'leg', 'tummy', etc.). These bout parameters were confirmed by two observers viewing raw video during pilot tests. Behavioural recordings and analysis were conducted by a genotype-blind experimenter. Given the 40 s minimum required to be classified as a bout, median data and cumulative probability were used to compare genotypes. Kolmogorov–Smirnov and the Wilcoxon rank-sum tests were used for statistical analysis. All individual data points are plotted with crossbar denoting median.

Fibre-photometry-based optical chloride measurements. Mice ($n = 6$ WT and 6 KO) were injected with 400 nl of AAV-hSyn-SuperClomeleon into the LGN (from Bregma A-P, -2 mm; M-L, -2.1 mm; D-V, -2.5 mm) and chronically implanted with optical fibres (400 μ m, Doric lenses) targeted directly above LGN (D-V, 2.1 mm). Following at least two weeks of virus expression, FRET-based measurement of visual-evoked $[Cl^-]$ responses was performed as previously described¹³. Briefly, CFP excitation was achieved through a fibre-coupled LED (Thorlabs) light source, filtered using a 434 nm clean-up filter (MF434-17 Thorlabs) and light was delivered to the LGN via a 600 μ m, 0.48 NA optic patch cord (Fig. 3a). SuperClomeleon CFP and YFP emissions were separated using a single-edge beam splitter (FF511-Di01, Semrock) and collected using a two femtowatt silicon photoreceiver (Newport). Signal was digitized and recorded using a TDT signal acquisition system (Tucker-Davis Technologies). Visual stimulation consisted of a 50 ms LED light pulses delivered to the eye contralateral to the recorded LGN. Normalized delta fluorescence ($\Delta F/F$) was calculated for evoked responses relative to the baseline fluorescence level before each event (1-s window) and smoothed with a convolution filter (50 ms half-width). The minimum signal within a 500 ms window following stimulation was considered the peak response. For pharmacological SK channel enhancement, mice were injected with 25 mg kg⁻¹ 1-EBIO following 30 min of baseline recordings. Maximal drug effects were estimated from evoked responses recorded between 30 min to 1 h following injection.

Visual detection task. Mice (C57, $n = 8$ WT and 9 KO from 2 independently tested cohorts; *Sst-Cre*, $n = 8$ WT and 8 KO from 2 independently-tested cohorts; 1-EBIO, 7 WT-Veh, $n = 7$ WT 1-EBIO, 7 KO veh., 7 KO 1-EBIO) were food restricted to 85–90% of their *ad libitum* body weight and training occurred in a custom-built test chamber as previously described³². During testing, a white noise indicated that a new trial was available, and mice had to continuously break an infrared barrier for 500–700 ms to initiate a trial. Upon successful initiation, a 50 ms visual stimulus was presented randomly either on the left or right side. Correct response at the corresponding nose-poke unit resulted in a milk reward (10 μ l evaporated milk, Nestle) that was available for 15 s. Following an inter-trial interval of 5 s, a new trial became available. Response at the incorrect location resulted in immediate blockage of poke access and a 30 s timeout before the next trial. To test for distractibility, in one third of the trials a 50 ms distractor appeared at the opposite location of where the stimulus would be displayed in the 200 ms time window before successful initiation. For drug treatment experiments, mice were injected with 1-EBIO (Tocris #1041; 25 mg kg⁻¹ in 10% DMSO; subcutaneous injection) or vehicle 30 min before testing. Wilcoxon rank-sum test was used for statistical analysis. All individual data points are plotted with crossbar denoting median.

Fura-2AM calcium imaging. Freshly prepared brain slices from P21–P28 male mice ($n = 4$ WT and 5 KO) were superfused with oxygenated ACSF (125 nM NaCl, 25 nM NaHCO₃, 25 nM glucose, 2.5 nM KCl, 1.25 nM NaH₂PO₄, 1.2 nM MgCl₂, 2 nM CaCl₂, 1.7 nM ascorbic acid) and recorded at 30–34 °C. Cells ($n = 37$ WT, 36 KO) were filled with Fura-2AM dye (Molecular Probes) using a Picospritzer II that applied ~ 10 pounds per square inch (psi) for 1 min through a 1–2 M Ω glass pipette. Slices were given 1 h to recover in ACSF before imaging. Individual cells were imaged using an Olympus BX61WI microscope with an attached CoolSnapF2 camera. Pixel intensity was measured using ImageJ software with whole-field background correction. Ratiometric values were converted to $[Ca^{2+}]_i$ using the following equation

$$[Ca^{2+}]_i = K_d(R - R_{min}) / (R_{max} - R) S_{D2} / S_{B2}$$

where K_d refers to the Ca^{2+} dissociation constant (140 nM), R refers to the ratiometric measurement of the observed cell, R_{max} and R_{min} correspond to the ratio under conditions of saturated Ca^{2+} levels and in zero Ca^{2+} , respectively. The values of S_{B2} (bound state) and S_{D2} (free Ca^{2+} state) are proportional to the fluorescence excited by 380 nm under conditions of saturated Ca^{2+} levels and in zero Ca^{2+} , respectively. Unpaired t -tests were used to compare nanomolar concentrations between the two groups. Data presented as mean \pm s.e.m.

Open field. Locomotor activity in mice (C57, $n = 30$ WT and 31 KO from 3 independently tested cohorts; mixture of C57 and 129 mice (referred to hereafter as mix), 10 WT and 11 KO from 1 cohort; amphetamine-treated, 5 WT veh., 5 WT amph., 6 KO veh., 6 KO amph.; 1-EBIO, 13 WT veh., 13 WT 1-EBIO, 13 KO veh., and 15 KO 1-EBIO from 2 independently-tested cohorts; *Sst-Cre*, 21 WT and 22 KO from 2 independently tested cohorts) was evaluated over a 60 min period in an automated Omnitech Digiscan apparatus (AccuScan Instruments). Locomotor activity was assessed as total distance travelled (m). Anxiety-like behaviour was defined by number of rearings and time spent in the centre as compared to time spent in the perimeter (thigmotaxis) of the open field. For drug treatment experiments, mice were placed in the open field arena for 30 min before amphetamine (3 mg kg⁻¹ in saline; intraperitoneal injection), 1-EBIO (25 mg kg⁻¹ in 10% DMSO; subcutaneous injection), or vehicle injections. Mice were then returned to open field arena for an additional 30–90 min. Two-way repeated measures ANOVA with Bonferroni post-hoc tests were used for statistical analysis. Time-binned data ('distance') presented as mean \pm s.e.m. Summated data ('total distance') plotted as individual data points with crossbar denoting mean. One C57 knockout mouse was excluded for escaping from the arena during testing.

Grooming. Young adult male mice (C57, $n = 9$ WT and 13 KO from 1 cohort; mix, $n = 10$ WT and 12 KO from 1 cohort) were used for analysis of grooming behaviour. Individually housed animals were habituated in the testing room for one hour before experimentation. Mice were video-taped for 2 h under 2 lx (red light) illumination. Grooming behaviours were coded from 19:00–21:00 (2 h beginning at the initiation of the dark cycle). This segment was analysed using Noldus Observer software and the total amount of time in the 2 h segment spent grooming was determined. A genotype-blind observer recorded all types of grooming, including incidences of face-wiping, scratching/rubbing of head and ears, and full-body grooming. Two-tailed t -tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

Rotarod. Motor coordination was assessed in mice (C57, $n = 19$ WT and 21 KO from 2 independently tested cohorts; mix, $n = 10$ WT and 10 KO from 1 cohort) using an accelerating rotarod test (Med Associates) over the course of two days. On the first day (training day) animals underwent three 5 min trials at a constant speed (16 r.p.m.). On the second day (testing day), animals underwent three 5 min trials at accelerating speeds (4–40 r.p.m.). For all trials, the latency to fall was determined. Animals were tested for three trials in a single day with an inter-trial interval of 10–30 min. Two-way repeated measures ANOVA with Bonferroni post-hoc tests were used for statistical analysis. Data plotted as mean \pm s.e.m. Two C57 knockout mice were excluded for jumping off the rotarod during testing.

Hanging wire. Mice (C57, $n = 12$ WT and 11 KO from 1 cohort; mix, 10 WT and 12 KO from 1 cohort; 1-EBIO, 6 WT veh., 6 WT 1-EBIO, 6 KO veh., and 6 KO 1-EBIO from 1 cohort; *Sst-Cre*, 12 WT and 11 KO from 1 cohort) were suspended 40 cm above the ground from a 2 mm horizontal wire. The average of three trials with an inter-trial interval of 5 min was recorded. For drug treatment experiments, mice were injected with 1-EBIO (Tocris #1041; 25 mg kg⁻¹ in 10% DMSO; subcutaneous injection) or vehicle 30 min before testing. Two-tailed t -tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

Gait. The forepaws of the mice (C57, $n = 10$ WT and 11 KO from 1 cohort) were painted green and the hindpaws were painted pink. After a two minute habituation trial, the mice were allowed to walk down a 50 cm track. The length and width of each stride were measured by an observer blind to genotype and the averages were recorded. Two-tailed t -tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

Hot plate. Mice (C57, $n = 20$ WT and 21 KO from 2 independently tested cohorts; mix, 10 WT and 12 KO from 1 cohort) were placed onto a heating block set to 55 °C surface temperature (Columbus Instruments). Latency to lick a forepaw or hindpaw was measured. The average of three trials with an inter-trial interval of 5 min was recorded. Two-tailed t -tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

Acoustic startle and pre-pulse inhibition. Auditory abilities and sensory motor function in mice (C57, $n = 20$ WT and 21 KO from 2 independently-tested cohorts; mix, 10 WT and 12 KO from 1 cohort) was measured using Hamilton Kinder Scientific Pre-pulse Startle Monitor chambers with Startle Monitor software. On the first day, mice underwent a 5 min acclimation trial in the acoustic startle boxes. On the second day, half of the mice were tested on the acoustic startle protocol while the other half was tested on the pre-pulse inhibition protocol. These protocol groups were switched on the third day. To test acoustic startle, mice are presented with pulses of various decibel levels without pre-pulses for approximately 30 min. The testing session is preceded by a 5 min exposure to 65 dB background noise. Each mouse then receives a total of 92 stimuli (trials) with inter-trial intervals ranging from 7–23 s presented in pseudo-random order. The stimuli include a presentation of eight pulse-alone trials (120 dB, 40 ms pulse, four at the beginning and

four at the end of the session), 77 pulse trials (seven each of 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, and 120 dB, 40 ms pulse), and seven trials each without pulse or pre-pulse presentation. To test pre-pulse inhibition, mice were once again exposed to 65 dB background for 5 min before testing. Each mouse received a total of 57 stimuli (trials) with inter-trial intervals ranging from 7–23 s presented in pseudorandom order. The stimuli include a presentation of eight pulse-alone trials (120 dB, 40 ms pulse, four at the beginning and four at the end of the session), 35 pre-pulse trials (seven each of 70, 75, 80, 85 and 90 dB, 20 ms pre-pulse given 100 ms before a 120 dB, 40 ms pulse), and seven trials each without pulse or pre-pulse presentation. The pre-pulse inhibition percentage within each test session was calculated as follows: $(100 - (\text{mean pre-pulse response} / \text{mean pulse response}) \times 100)$. For all experiments, response to startle stimuli is measured in newtons (N). Startle at each pulse level is averaged across trials and then across animals in a treatment group. Two-way repeated measures ANOVA with Bonferroni post-hoc tests were used for statistical analysis. Data plotted as mean \pm s.e.m.

Three-chambered social interaction assay. Littermate male mice (C57, $n = 10$ WT and 11 KO from 1 cohort; mix, 10 WT and 12 KO from 1 cohort) were used for all tests. Age and size-matched 129 male target subjects ('stranger 1' and 'stranger 2') were habituated to being placed inside wire cages for three days (20 min per day) before beginning of testing. Test mice were habituated to the testing room for at least 1 h before the start of behavioural tasks. The social test apparatus consisted of a transparent acrylic box with removable floor and partitions dividing the box into three chambers. The wire cages used to contain the stranger mice were cylindrical, 11 cm in height, a bottom diameter of 10.5 cm with the bars spaced 1 cm apart (Galaxy Cup, Spectrum Diversified Designs). An inverted metal can was placed on the top of the cage to prevent the test mice from climbing on the top of the wire cage. For the sociability test, the test animal was introduced to the middle chamber and left to habituate for 10 min. Following this period, the middle chamber doors were opened and the test mouse was allowed to freely explore all three chambers for an additional 10 min. The test mouse was then returned to the middle chamber, after which an unfamiliar mouse (stranger 1) was introduced into a wire cage on one of the side-chambers and an empty wire cage (E) on the other side-chamber. The dividers were then raised and the test animal was allowed to freely explore all three chambers over a 10 min session. Following the 10 min session, the mouse was returned to the middle chamber while a novel stranger mouse (stranger 2) was inserted in the wire cage, previously empty, and again the test animal was left to explore for a 10 min session. The time spent by the mouse (nose-point) in close proximity (~ 5 cm) to the wire cages was calculated using automated software (Noldus Ethovision 9). The release of the animals and relative position of social and inanimate targets was counterbalanced. However, for each individual test animal the location of stranger 1 was maintained during stranger-1-E and stranger-1-stranger-2 testing of the social behaviour. One-way ANOVA with Bonferroni multiple comparison tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

Resident intruder. Mice (C57, $n = 10$ WT and 10 KO from 1 cohort; 1-EBIO, 6 WT veh., 6 WT 1-EBIO, 6 KO veh., 6 KO 1-EBIO from 1 cohort; *Sst-Cre*, 6 WT and 6 KO from 1 cohort) were individually housed for 2 weeks before testing with bedding left undisturbed for one week before testing. Animals were tested in a room with 10lx lighting. On the test day, animals were habituated to the test room for 1 h before testing. After acclimation, an age- and weight-matched conspecific male stranger mouse was introduced to the home cage of the test animal. The subsequent interactions were videotaped for 5 min. All interactions were analysed using Noldus Observer by a genotype-blind observer. Aggressive interactions were defined by instances of biting, fighting, and tail-rattling. For drug treatment experiments, mice were injected with 1-EBIO (Tocris #1041; 25 mg kg⁻¹ in 10% DMSO; subcutaneous injection) or vehicle 30 min before testing. Wilcoxon rank-sum and

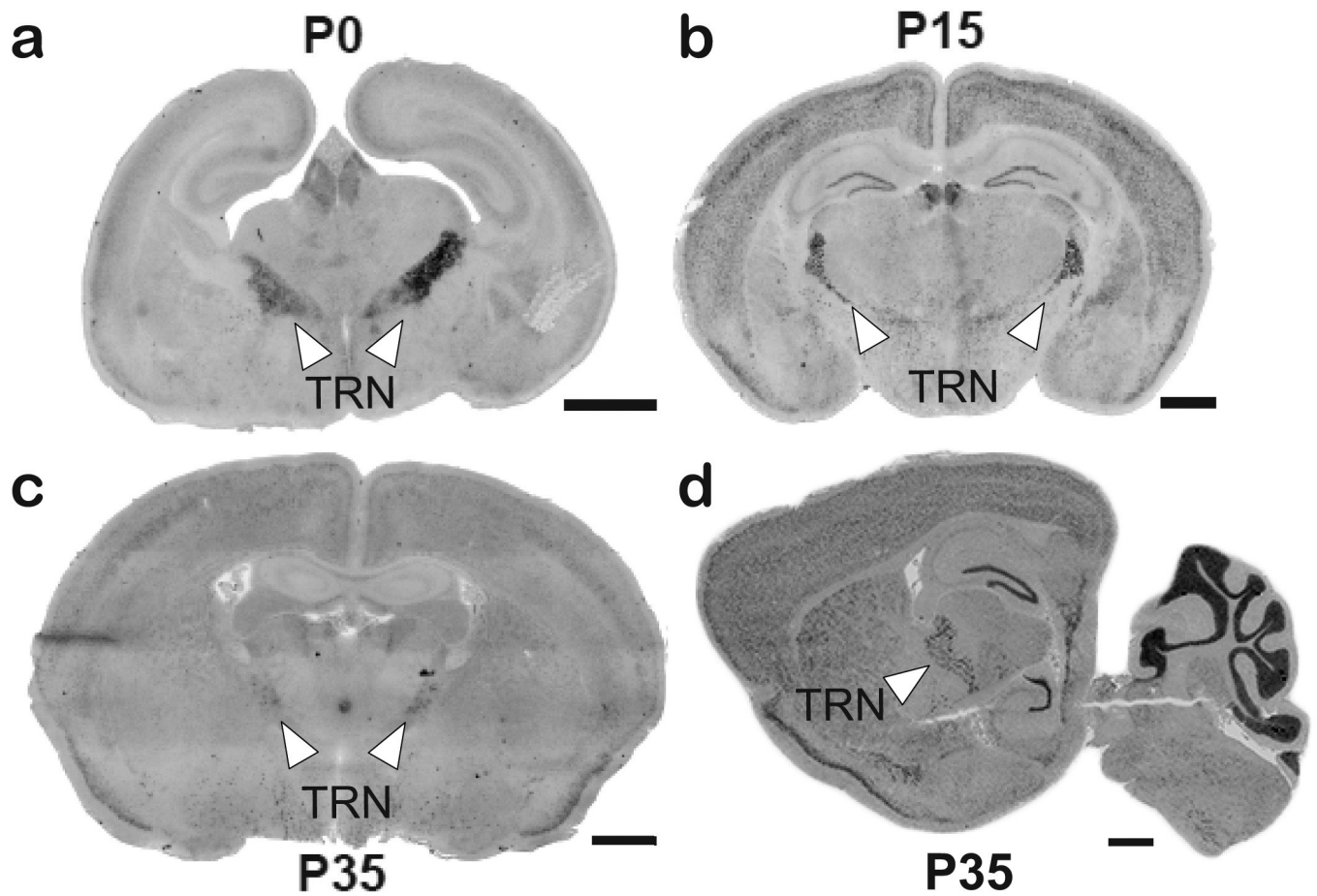
Kruskal–Wallis non-parametric tests were used for statistical analysis because wild-type mice did not pass normality tests. All individual data points are plotted with crossbar denoting median.

Inhibitory avoidance. Mice (C57, $n = 24$ WT and 23 KO from 2 independently-tested cohorts; mix, 10 WT and 12 KO from 1 cohort, 1-EBIO, 7 WT veh., 8 WT 1-EBIO, 9 KO veh., and 10 KO 1-EBIO from 2 independently-tested cohorts; *Sst-Cre*, 11 WT and 12 KO from 1 cohort) underwent one training session followed by 24 h and 48 h post-training trials. On the training day, mice were placed in the light side of the inhibitory avoidance box (Ugo Basile) and allowed to habituate for 30 s. To quantify pre-shock latency and the effects of baseline locomotor activity levels on this task, the door to the dark side of the box was then opened and the latency to cross was measured. Upon entering the dark side of the box, the mice were given a 0.5 mA shock for 2–4 s, followed by a 60 s post-shock habituation. On the probe trials, mice were placed in the light side of the box with the door to the dark side already opened. Latency to cross was once again measured with a maximum duration of 9 min. For drug treatment experiments, mice were injected with 1-EBIO (Tocris #1041; 25 mg kg⁻¹ in 10% DMSO; subcutaneous injection) or vehicle 30 min before the shock training protocol. Two-way repeated measures ANOVA with Bonferroni post-hoc tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

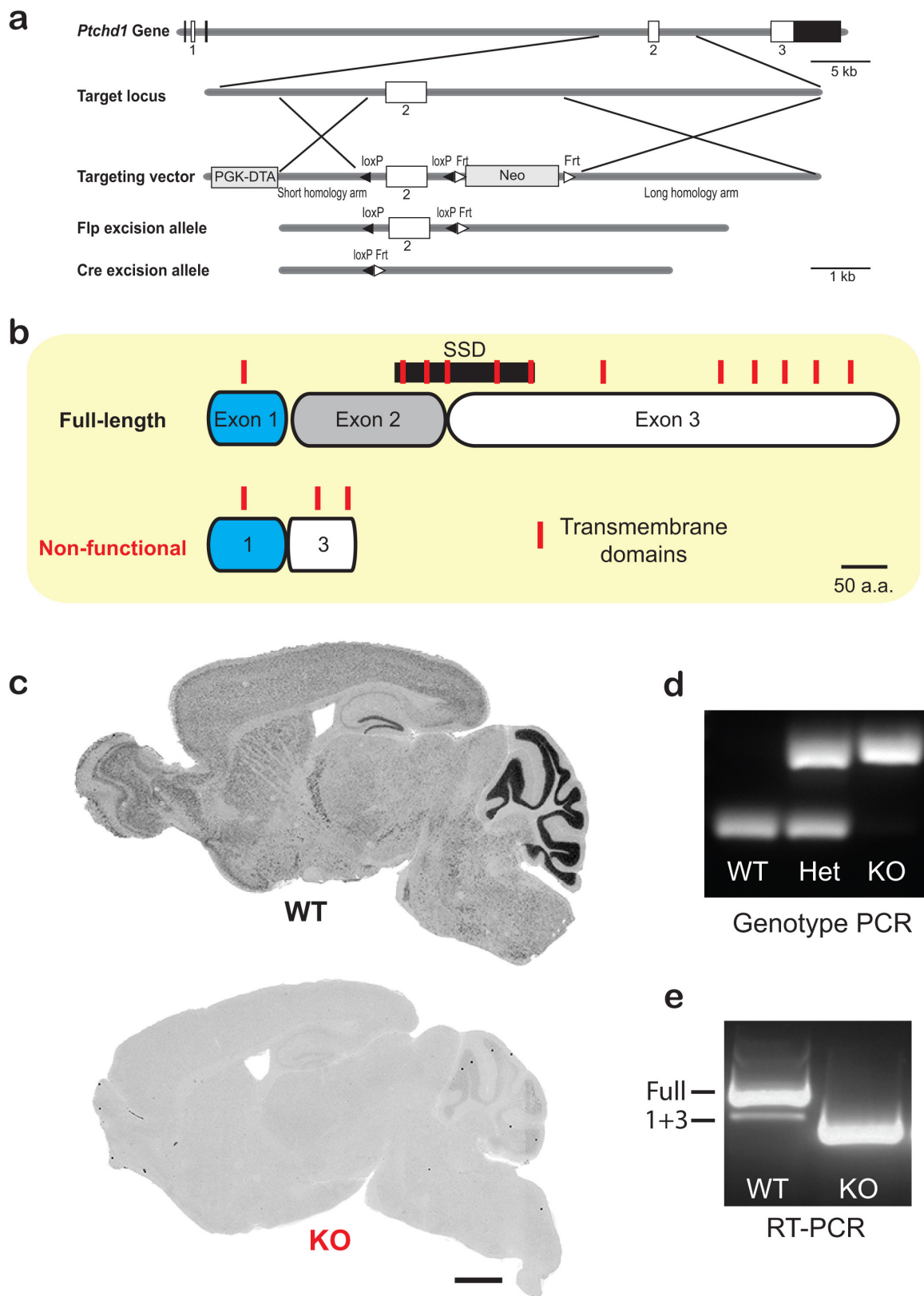
Fear conditioning. Mice (C57, $n = 10$ WT and 11 KO from 1 cohort) were tested for fear-induced freezing using Med Associates fear-conditioning chambers encased in sound attenuating cubicles optimized for near infrared (NIR) video recording. The training protocol involved a 3 min habituation period in the conditioning box, followed by three rounds of a 30 s tone +2 s of 0.75 mA shock +90 s rest. The final shock was followed by a 2 min post-training habituation. The following day, mice were returned to the conditioning box and the time spent freezing was measured. Four hours after, the mice were returned to the conditioning boxes with modifications. A white triangular insert was added to change the dimensions of the box and 0.1% acetic acid was sprayed onto the base of the box in order to change the scent of the box. Once again, the time spent freezing over the course of 5 min was measured. Freezing was measured using Video Freeze software package analysis of NIR recordings. Two-tailed *t*-tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

Morris water maze. Spatial learning testing in mice (C57, $n = 10$ WT and 10 KO from 1 cohort) was conducted using a testing pool that was 120 cm in diameter and a platform 8 cm in diameter. The platform was submerged 1 cm below the water surface. Pool water was maintained at $23.0 \pm 0.5^\circ\text{C}$ and made opaque by mixing in non-toxic white paint. During training, 90 s duration trials were used. If the animals did not find the platform within 90 s, the experimenter guided the animal to the platform. After reaching the platform, the animals were left for 15 s on top of the platform before being removed. Trials were administered for 5–6 days with four trials per animal per day with the platform located in the north-west (NW) quadrant. For two consecutive days after the training protocol, 60 s probe trials were performed (one per day). The reversal training commenced with the platform in the south-east quadrant, and proceeded as described above. The experimenter followed the animals' progress using tracking software outside of the testing room. Tracking and analysis were performed using the Noldus Ethovision software. Two-way repeated measures ANOVA with Bonferroni post-hoc tests and one-way ANOVA with Bonferroni multiple comparison tests were used for statistical analysis. All individual data points are plotted with crossbar denoting mean.

50. Fisher, S. P. et al. Rapid assessment of sleep-wake behavior in mice. *J. Biol. Rhythms* **27**, 48–58 (2012).



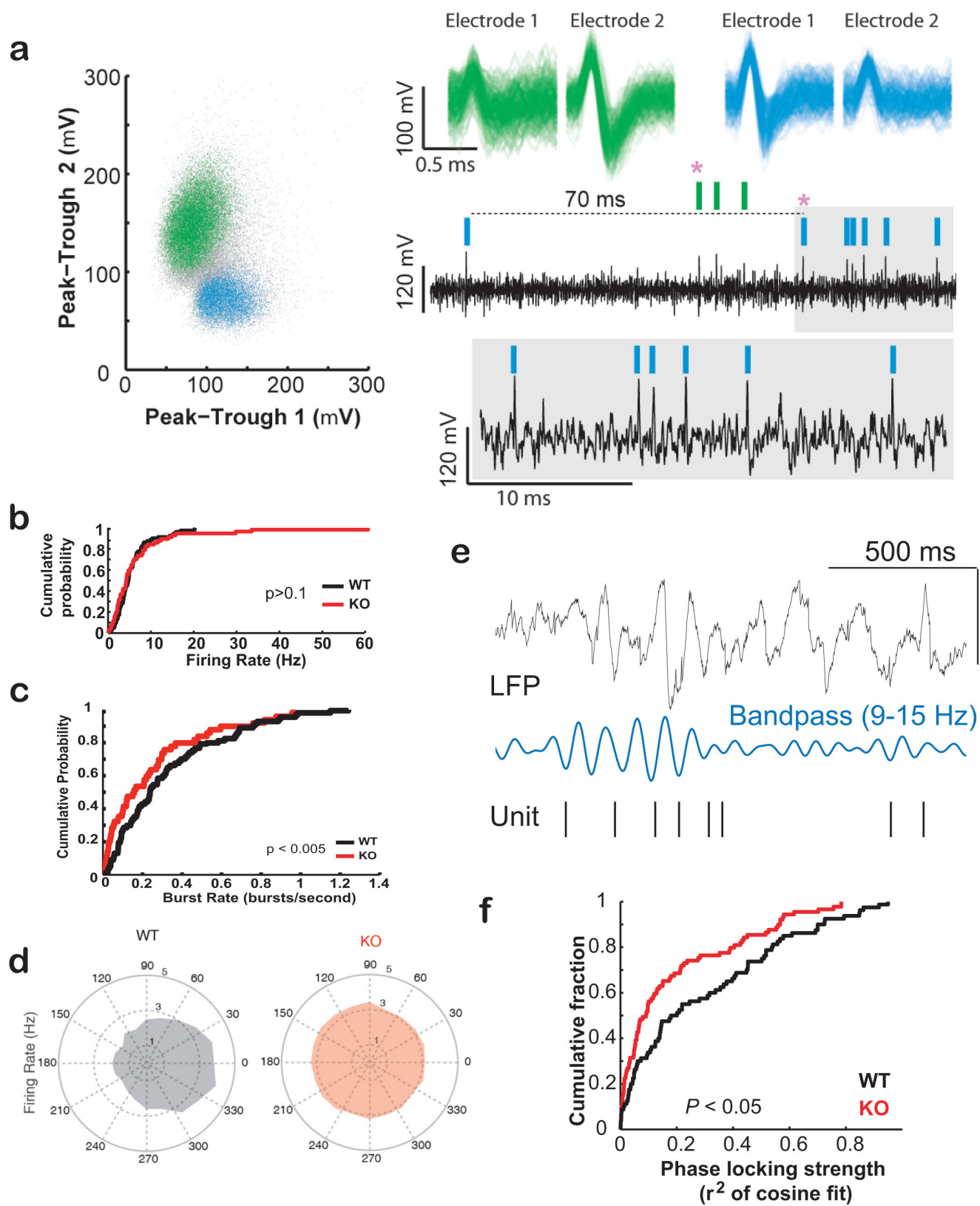
Extended Data Figure 1 | Developmental expression pattern of *Ptchd1*. a–d, *In situ* hybridization labelling of *Ptchd1* mRNA at P0 (coronal) (a), P15 (coronal) (b), and P35 (coronal and sagittal) (c–d) from 3 C57/Bl6 wild-type mice per age. White arrows indicate location of TRN region (scale bars, 1 mm).



Extended Data Figure 2 | Generation of *Ptchd1*-knockout mouse.

a, Schematic describing strategy to create *Ptchd1*-knockout mouse. Mice containing targeted allele were crossed to β -actin Flp mice to remove the Neo cassette and β -actin Cre mice to excise exon 2. **b**, Diagram depicting the 'full-length' and non-functional 'exons 1+3' *Ptchd1* isoforms. Genetic ablation of Exon 2 results in the removal of a majority of the

transmembrane domains normally present in the endogenous full-length isoform. **c**, *In situ* hybridization probes targeting exon 2 confirm successful genetic ablation of full-length *Ptchd1* mRNA (scale bar, 1 mm). **d**, PCR genotyping confirms deletion of exon 2 from genome of male knockout mice. **e**, qPCR of wild-type and knockout cDNA samples shows removal of full-length *Ptchd1* isoform. Het, heterozygous.

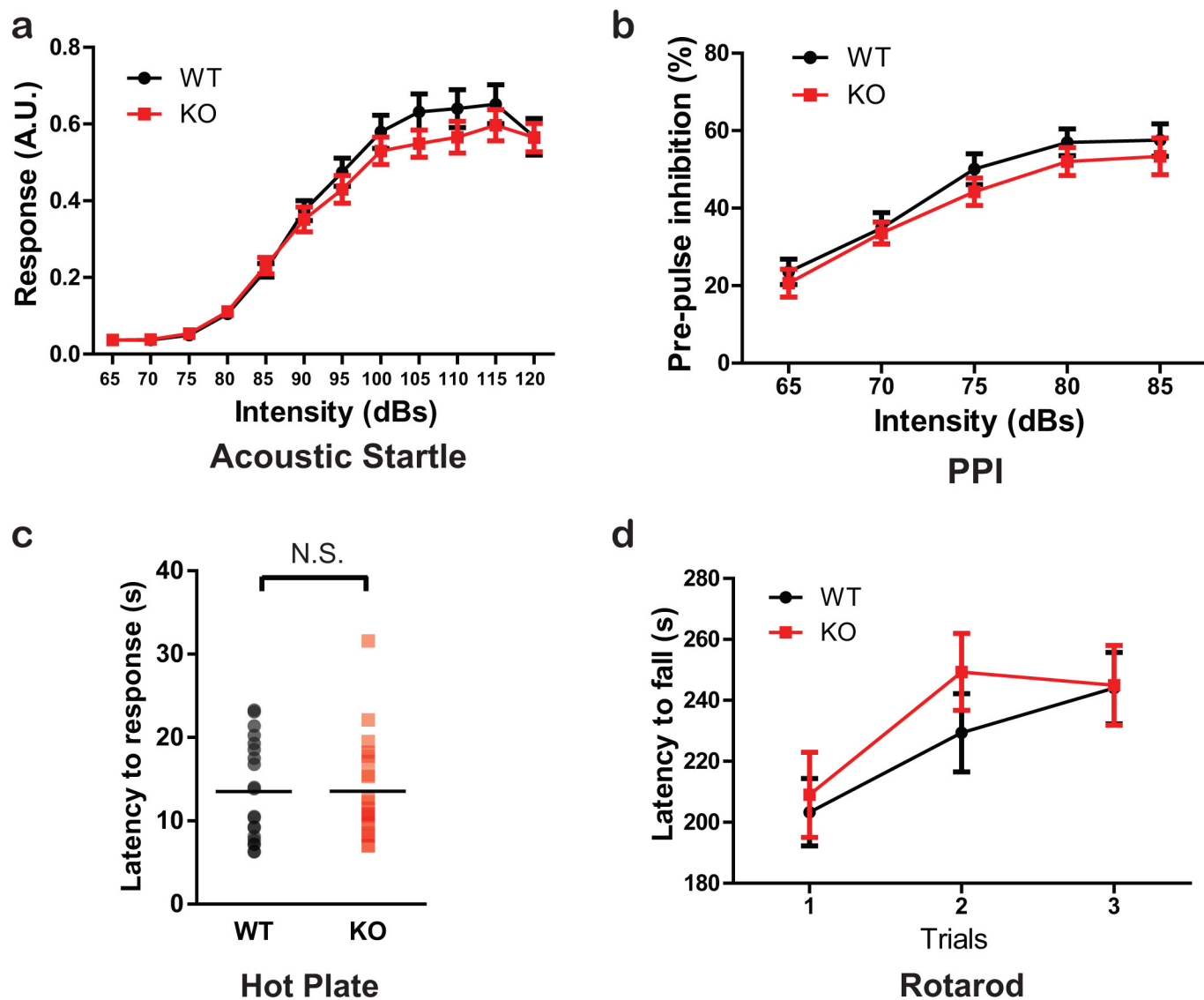


Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Burst and spindle phase-locking characteristics of *Ptchd1*-knockout and wild-type TRN neurons *in vivo*.

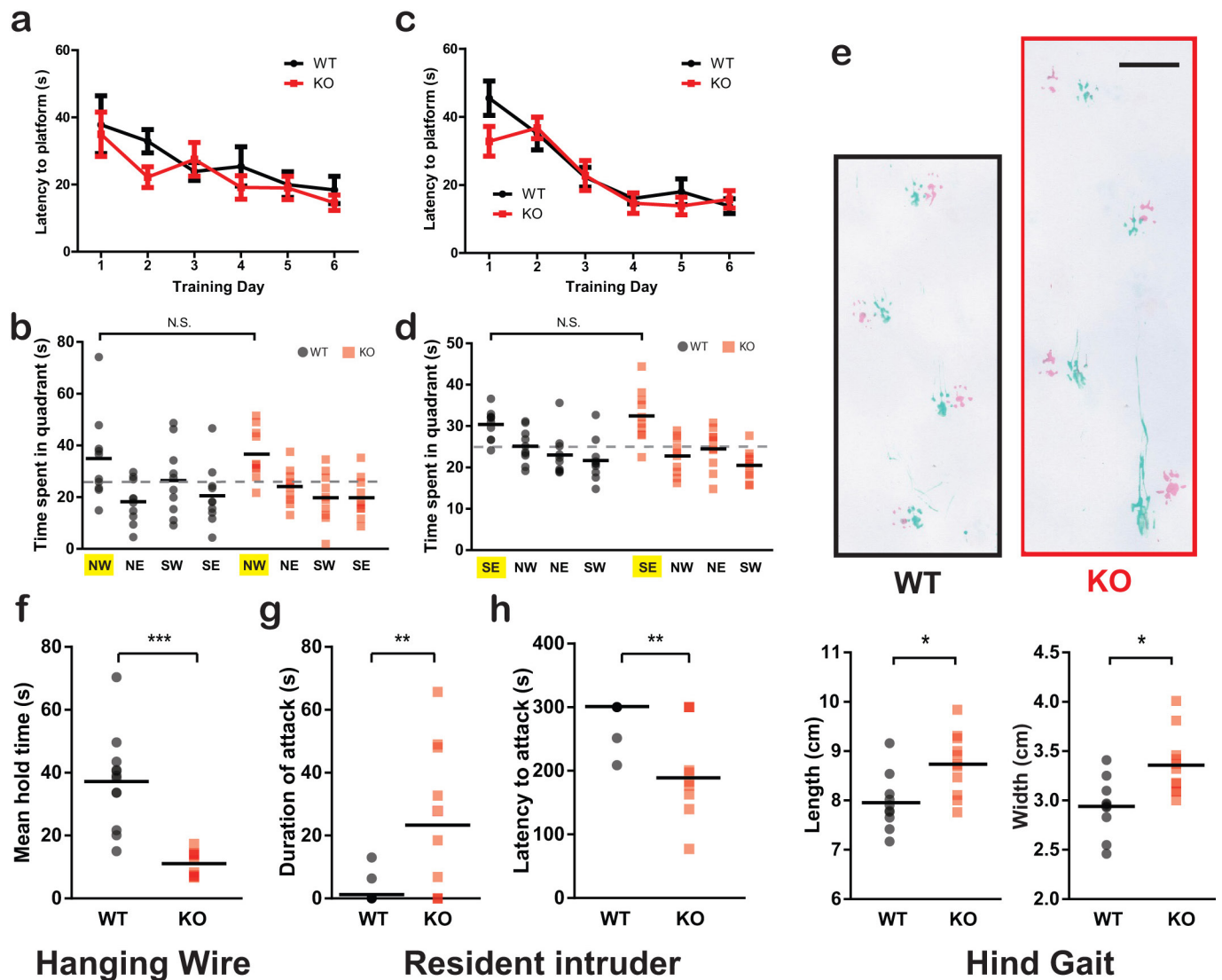
a, Left, example of unit clustering for a stereotrode recording. Two units (green and blue) are clearly separated when plotting peak-trough of the two electrodes of the stereotrode against each other. Right, spike-wave form of the two clustered units as they appear on the two electrodes of the stereotrode. Raw trace below shows a burst discharge (asterisk) of each unit during non-rapid eye movement sleep with coloured ticks indicating corresponding individual spikes. A burst was identified as at least 2 spikes with an inter-spike interval of ≤ 10 ms, preceded by a period of 70 ms silence. Enlarged trace shows the accelerando-decelerando firing pattern characteristic for a TRN burst. **b**, Firing rate during non-rapid eye movement sleep is comparable between genotypes ($n = 89$ WT,

80 KO cells from 4 WT, 3 KO mice; $P > 0.1$, Kolmogorov-Smirnov test). **c**, *Ptchd1*-knockout TRN neurons show reduced propensity to generate bursts, even when excluding the 10% of knockout cells with the highest firing rate ($n = 89$ WT, 72 KO cells from 4 WT, 3 KO mice; $P < 0.05$, Kolmogorov-Smirnov test). **d**, Spindle-phase histogram for an example wild-type and knockout neuron. Note, the wild-type neuron shows a preferred phase around the peak (0 degrees) of the spindle oscillation in wild type but not knockout. **e**, Example local field potential recording (LFP; top) showing the temporal alignment of TRN spikes (bottom) to the preferred phase of the spindle activity (9–15 Hz, middle). **f**, *Ptchd1*-knockout mice show reduced phase-locking strength to spindle activity compared to wild-type littermates ($n = 89$ WT, 80 KO cells from 4 WT, 3 KO mice).



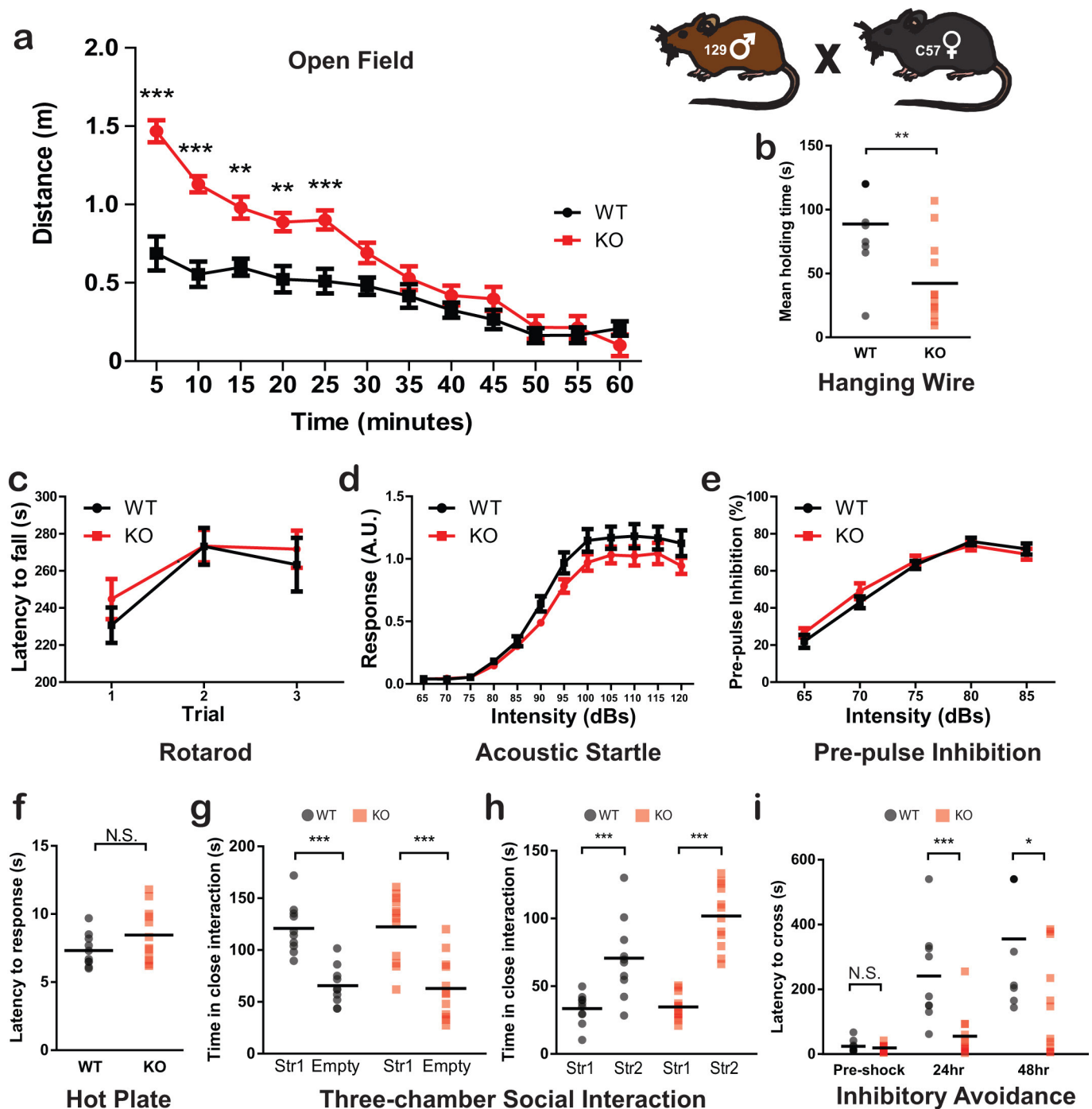
Extended Data Figure 4 | *Ptchd1*-knockout mice have intact sensory responses and rotarod performance. a–c, Normal acoustic startle (a), pre-pulse inhibition (PPI; b) and hot plate response (c) in *Ptchd1*-knockout mice ($n = 20$ WT (a–c), 20 KO (b), 21 KO (c)). d, *Ptchd1*-

knockout mice show normal motor coordination on the accelerating rotarod test ($n = 19$ WT, 20 KO). Two-tailed t -tests (c) and two-way repeated measures ANOVA with Bonferroni post-hoc tests (a–b, d) were used for statistical analysis. Error bars, mean \pm s.e.m. NS, not significant.



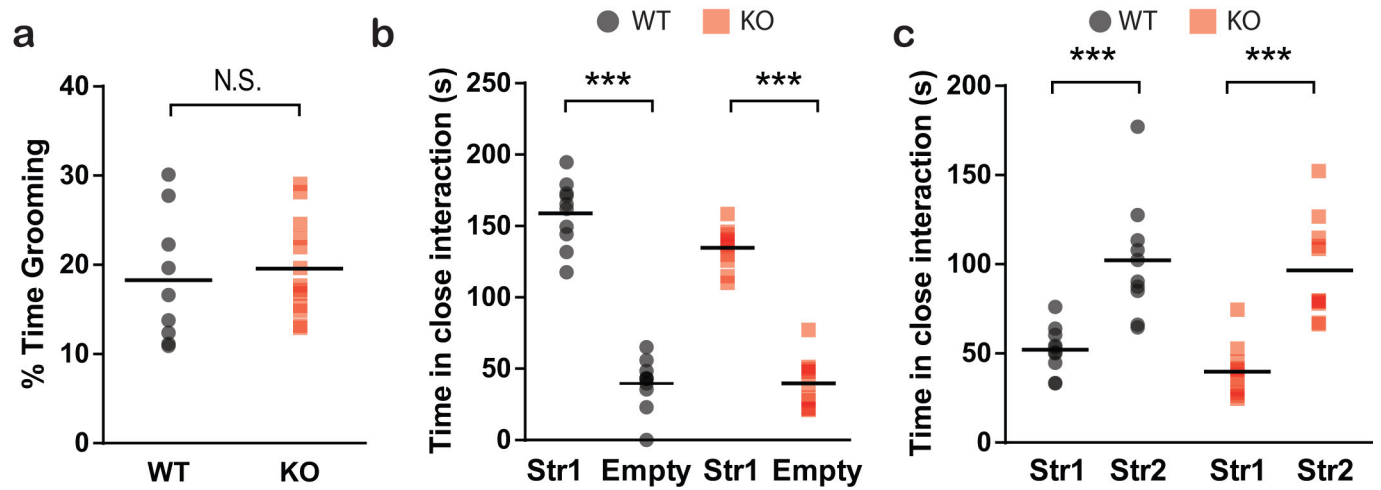
Extended Data Figure 5 | Intact spatial learning but motor and aggression abnormalities in *Ptchd1*-knockout mice. **a**, Comparable learning curves between wild-type and knockout mice during cued training protocol. **b**, Intact spatial learning demonstrated in 24 h probe trial. **c**, *Ptchd1*-knockout mice show normal reversal learning curve. **d**, No significant difference between wild-type and knockout mice in 24 h probe trial after reversal learning protocol ($n = 10$ WT, 10 KO). **e**, Representative images of wild-type (black) and knockout (red) strides. Forepaw position is represented by green paint and hindpaw position is represented by pink paint (scale bar, 2 cm). Quantification reveals elongated stride length and width ($n = 10$ WT, 11 KO). **f**, Knockout mice show marked reductions in

grip strength as measured by the hanging wire test ($n = 12$ WT, 11 KO). **g**, **h**, Knockout mice attack intruder mice for a longer duration (**g**) and with a shorter latency to attack (**h**) in the resident-intruder test for aggression ($n = 10$ WT, 10 KO). Two-way repeated measures ANOVA with Bonferroni post-hoc tests (**a**, **c**), one-way ANOVA with Bonferroni multiple comparison tests (**b**, **d**), two-tailed *t*-tests (**e**, **f**) and Wilcoxon rank-sum tests (**g**, **h**) were used for statistical analysis. Chance performance (25%) represented by dashed grey lines (**b**, **d**). Error bars, mean \pm s.e.m (**a**, **e**); mean (**b**, **d**, **e**-**f**), median (**g**, **h**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.



Extended Data Figure 6 | Hyperactivity, hypotonia, and learning deficits in C57/129 *Ptchd1*-knockout mice. **a**, *Ptchd1*-knockout mice showed increased locomotor activity ($n = 10$ WT, 11 KO). **b**, **c**, Knockout mice show decreased mean holding time in the hanging wire test (**b**) ($n = 10$ WT, 12 KO), but normal motor coordination in the rotarod task (**c**) ($n = 10$ WT, 10 KO). **d**–**f**, Sensory responses as measured by acoustic startle (**d**), pre-pulse inhibition (**e**), and hot plate (**f**) are also normal in knockout mice ($n = 10$ WT, 12 KO). **g**–**h**, Normal sociability (**g**) and novel

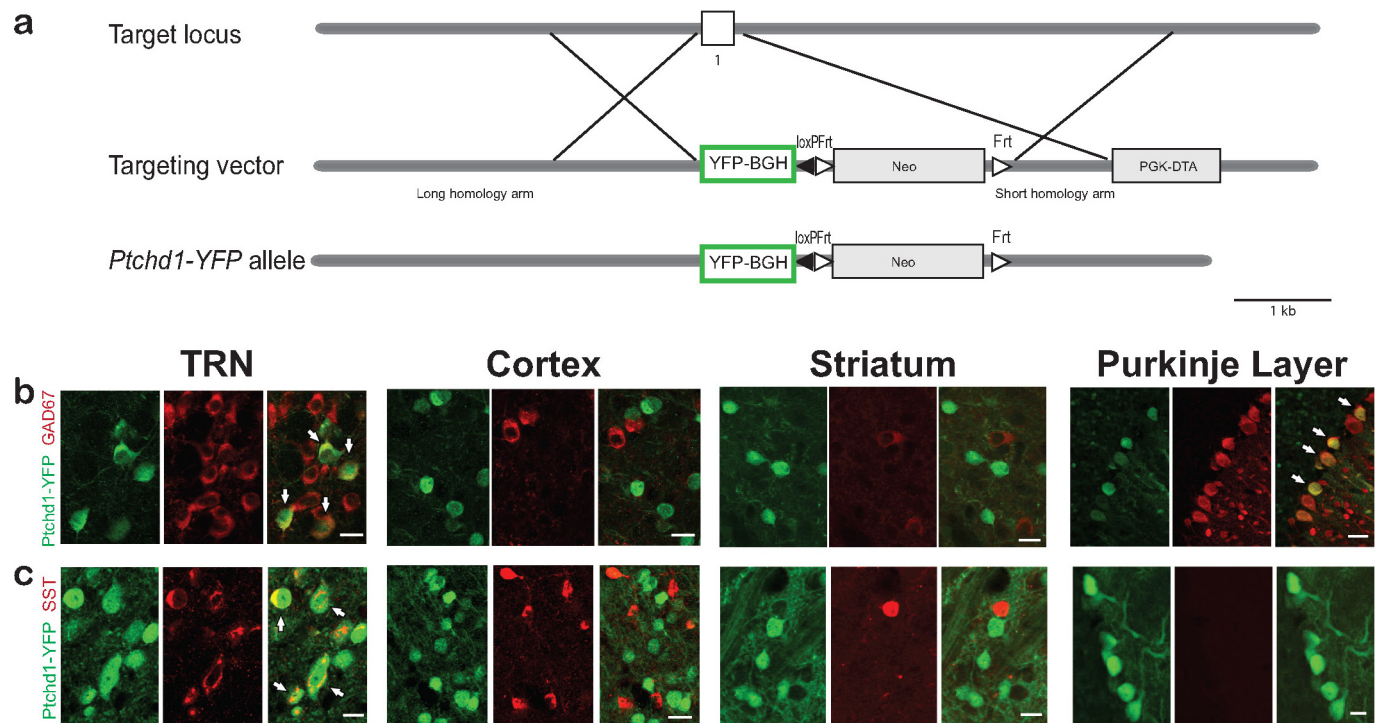
social recognition (**h**) in mixed background *Ptchd1*-knockout mice ($n = 10$ WT, 11 KO). **i**, Knockout mice show impaired associative learning and memory in the inhibitory avoidance task ($n = 9$ WT, 12 KO). Two-tailed t -tests (**b**, **f**), one-way ANOVA with Bonferroni multiple comparison tests (**g**, **h**), and two-way repeated measures ANOVA with Bonferroni post-hoc tests (**a**, **c**–**e**, **i**) were used for statistical analysis. Error bars, mean \pm s.e.m.; horizontal bars, mean (**b**, **f**–**i**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.



Three-chamber Social Interaction

Extended Data Figure 7 | Normal grooming and social interaction behaviours in *Ptchd1*-knockout mice. **a**, Knockout mice do not show excessive or injurious grooming behaviours ($n = 9$ WT, 13 KO). **b**, **c**, Knockout mice spent comparable amounts of time interacting with stranger mice in the three-chambered social interaction task (**b**) and

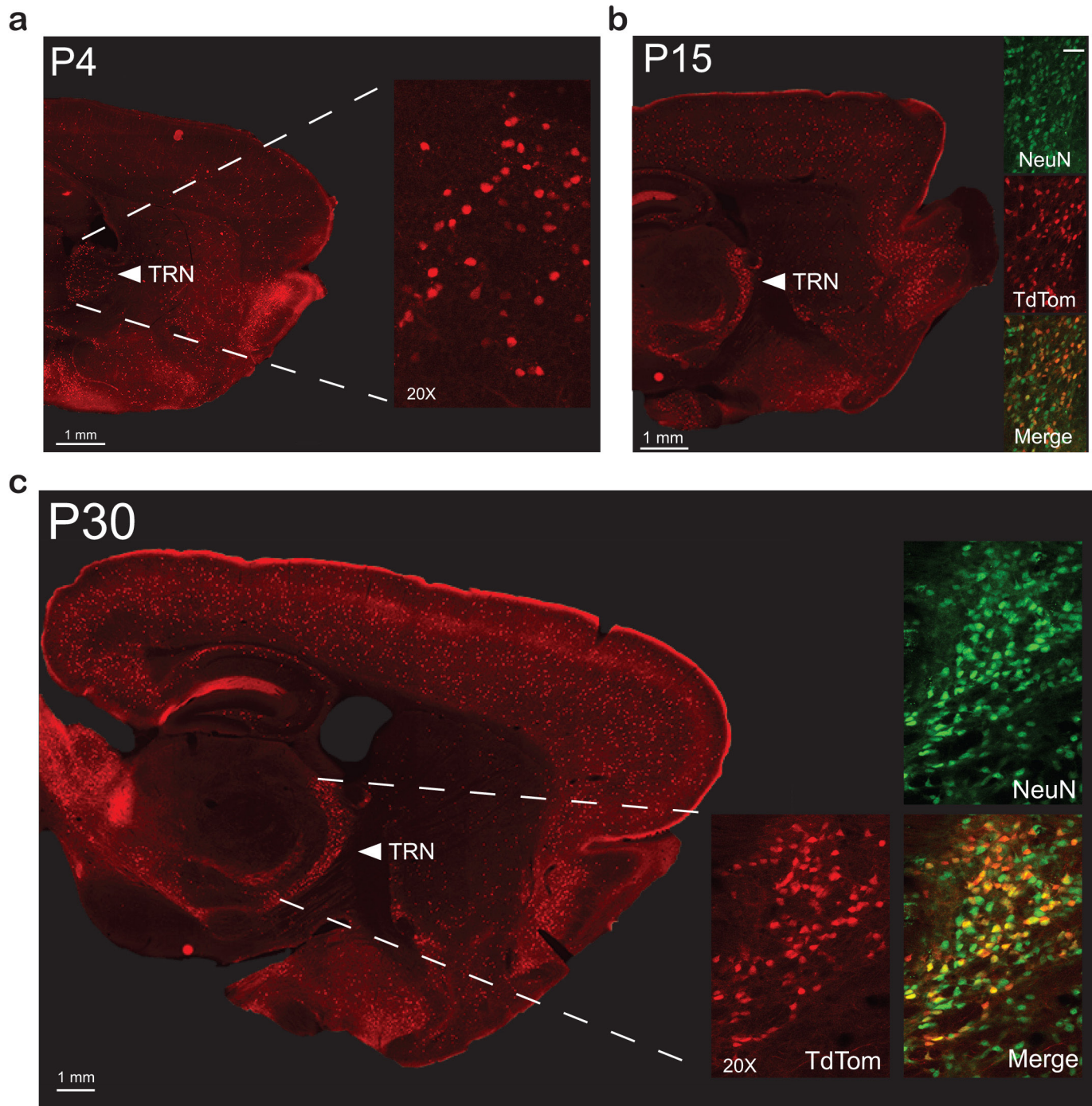
display normal social novelty behaviours (**c**) ($n = 10$ WT, 11 KO). Two-tailed t -tests (**a**) and two-way repeated measures ANOVA with Bonferroni post-hoc tests (**b**, **c**) were used for statistical analysis. Horizontal bars, mean. *** $P < 0.001$.



Extended Data Figure 8 | YFP overlap with somatostatin interneuron marker is primarily confined to the TRN in *Ptchd1*-YFP mice.

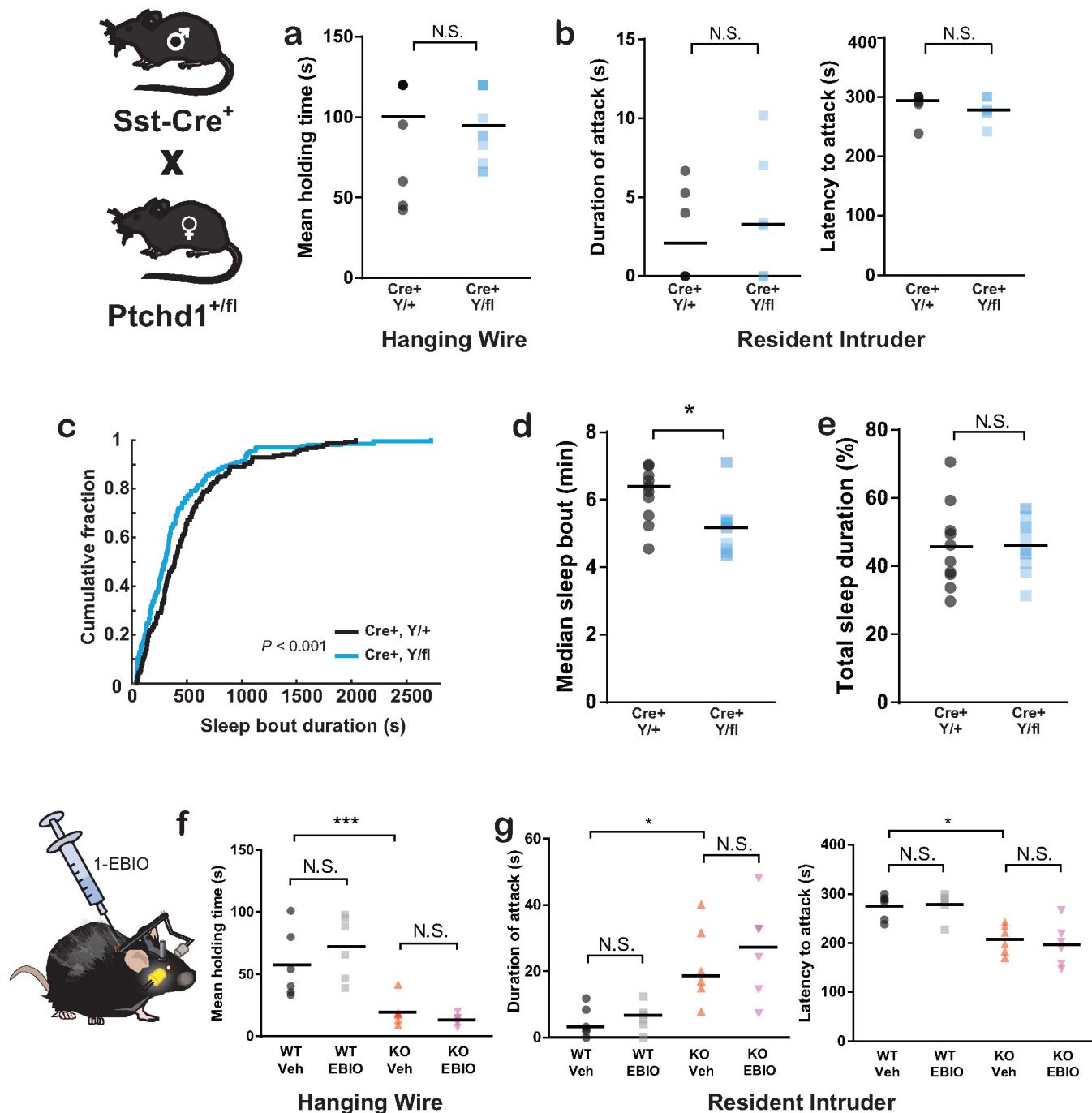
a, Schematic describing strategy to create *Ptchd1*-YFP mouse in which exon 1 was replaced with a YFP-bovine growth hormone poly-A tail

(BGH) cassette. **b**, YFP⁺ cells co-label with anti-GAD67 antibody in TRN and the Purkinje layer of the cerebellum, but not in cortex or striatum. **c**, YFP⁺ cells also co-label with anti-somatostatin antibody in TRN, but not in other structures. Arrows denote overlap. Scale bars, 20 μm.



Extended Data Figure 9 | *Sst*-Cre recombinase activity is early and robust in TRN neurons. **a**, The progeny of *Sst*-Cre⁺ mice crossed to mice showing Cre-dependent expression of the TdTomato fluorescent protein (*Sst*-Cre⁺ TdTomato⁺) show TdTomato⁺ cells in the TRN at P4. Inset

shows magnified image taken with 20× objective. **b**, **c**, At P15 (**b**) and P30 (**c**), Cre recombinase activity in the TRN of the *Sst*-Cre⁺ TdTomato⁺ mice brains is robust, as shown by the inset depicting the significant TdTomato overlap with the pan-neuronal marker NeuN.



Extended Data Figure 10 | Genetic disruption of *Ptchd1* TRN expression affects sleep stability but not grip strength or aggressive behaviours. **a, b**, *Sst-Cre⁺ Ptchd1^{Y/f}* mice appear normal in the hanging wire (**a**) ($n = 12$ *Ptchd1^{Y/+}*, 11 *Ptchd1^{Y/f}*) and resident intruder task (**b**) ($n = 6$ *Ptchd1^{Y/+}*, 6 *Ptchd1^{Y/f}*). **c–e**, *Sst-Cre⁺ Ptchd1^{Y/f}* mice show reductions in sleep bout duration as shown in cumulative probability plot and comparison of medians (**d**) with no differences in total time spent sleeping when compared to *Sst-Cre⁺:Ptchd1^{Y/+}* littermates (**e**)

($n = 10$ *Ptchd1^{Y/+}*, 10 *Ptchd1^{Y/f}*). **f, g**, 1-EBIO treatment has no effect on performance on the hanging wire (**f**) or resident intruder task (**g**) ($n = 6$ WT veh., 6 WT 1-EBIO, 6 KO veh., 6 KO 1-EBIO). Kolmogorov–Smirnov test (**a**), Wilcoxon rank-sum tests (**b, c**), two-tailed *t*-tests (**d**), and two-way repeated measures ANOVA with Bonferroni post-hoc tests (**f**), and Kruskal–Wallis with Dunn’s multiple comparisons tests were used for statistical analysis. Horizontal bars, median (**b, c, e–g**), mean (**d, f**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Candidalysin is a fungal peptide toxin critical for mucosal infection

David L. Moyes^{1*}, Duncan Wilson^{2†*}, Jonathan P. Richardson^{1*}, Selene Mogavero^{2*}, Shirley X. Tang¹, Julia Wernecke^{3,4}, Sarah Höfs², Remi L. Gratacap⁵, Jon Robbins⁶, Manohursingh Runglall^{1†}, Celia Murciano^{1†}, Mariana Blagojevic¹, Selvam Thavaraj¹, Toni M. Förster², Betty Hebecker^{2,7}, Lydia Kasper², Gema Vizcay⁸, Simona I. Iancu¹, Nessim Kichik^{1,9}, Antje Häder¹⁰, Oliver Kurzai¹⁰, Ting Luo¹¹, Thomas Krüger¹¹, Olaf Kniemeyer¹¹, Ernesto Cota⁹, Oliver Bader¹², Robert T. Wheeler⁵, Thomas Gutschmann³, Bernhard Hube^{2,13,14} & Julian R. Naglik¹

Cytolytic proteins and peptide toxins are classical virulence factors of several bacterial pathogens which disrupt epithelial barrier function, damage cells and activate or modulate host immune responses. Such toxins have not been identified previously in human pathogenic fungi. Here we identify the first, to our knowledge, fungal cytolitic peptide toxin in the opportunistic pathogen *Candida albicans*. This secreted toxin directly damages epithelial membranes, triggers a danger response signalling pathway and activates epithelial immunity. Membrane permeabilization is enhanced by a positive charge at the carboxy terminus of the peptide, which triggers an inward current concomitant with calcium influx. *C. albicans* strains lacking this toxin do not activate or damage epithelial cells and are avirulent in animal models of mucosal infection. We propose the name ‘Candidalysin’ for this cytolitic peptide toxin; a newly identified, critical molecular determinant of epithelial damage and host recognition of the clinically important fungus, *C. albicans*.

The ability of mucosal surfaces to discriminate between commensal and pathogenic microbes is essential to human health. The fungus *Candida albicans* is normally a benign member of the human microbiota, but is also responsible for millions of mucosal infections each year in immunocompromised hosts, often with severe morbidity¹. A defining feature of *C. albicans* pathogenesis is the transition from yeast to invasive filamentous hyphae². Hyphae damage mucosal epithelia and induce activation of the activating protein-1 (AP-1) transcription factor, c-Fos (via p38-MAPK) and the MAPK phosphatase MKP1 (via ERK1/2-MAPK), which trigger pro-inflammatory cytokine responses^{3–7}. These signalling events constitute a ‘danger response’ against invasive hyphae, thus serving as a sensor of pathogenic *C. albicans* invasion^{8–14}. However, it is unclear how *C. albicans* hyphae induce epithelial inflammatory responses and cell damage during mucosal infections. Here we identify and characterize Candidalysin, the first, to our knowledge, cytolitic peptide toxin isolated from any human fungal pathogen, as the hyphal factor critical for epithelial immune activation and *C. albicans* mucosal infection.

Ecelp is vital for epithelial activation and damage

Despite the well-known association between filamentation and virulence, the molecular mechanism underlying hypha-driven epithelial activation and mucosal damage has remained obscure. To elucidate this mechanism, we screened a panel of *C. albicans* gene deletion mutants that targeted key processes, pathways and proteins known or predicted to be associated with the yeast–hyphal transition and pathogenicity (62 strains). Only hypha-producing strains induced MKP1 phosphorylation

(p-MKP1), c-Fos, cytokines (IL-1 α , IL-6, G-CSF) and damage in oral epithelial cells (Extended Data Table 1). However, one *C. albicans* mutant (*ece1 Δ /*)¹⁵ formed normal hyphae but was incapable of inducing these epithelial danger responses. The *C. albicans* extent of cell elongation 1 gene (*ECE1*) is highly expressed by hyphae during epithelial infection (Extended Data Fig. 1a, b), and is predicted to encode a secreted protein¹⁶. To probe its function, we generated a panel of *C. albicans* *ECE1* mutants (Extended Data Table 2). The *ece1 Δ /* strain formed normal hyphae on (Extended Data Fig. 1c), and adhered to and invaded human epithelial cells similarly to wild-type *C. albicans* (Extended Data Fig. 1d, e). The strain *ece1 Δ /* was capable of extensive epithelial invasion, penetrating through multiple epithelial cells (Extended Data Fig. 1f). Despite this, invasive *ece1 Δ /* hyphae did not damage epithelia or induce p-MKP1/c-Fos-mediated danger responses or cytokine secretion (Fig. 1a–d). Thus, Ecelp is critical for epithelial damage and innate recognition of *C. albicans* hyphae *in vitro*.

Ecelp is critical for mucosal pathogenesis

We next assessed the role of *ECE1* in two *in vivo* models of *C. albicans* mucosal infection. In murine oropharyngeal candidiasis (OPC)¹⁷, mice infected with *C. albicans* wild-type or *ECE1* re-integrant (*ece1 Δ /* + *ECE1*) strains exhibited disease symptoms, including extensive hyphal invasion of the tongue epithelium, micro-abscesses of infiltrating neutrophils and tissue damage (Fig. 1e, f, h, i). In contrast, tongue tissue from *ece1 Δ /*-infected animals ($n = 17/20$) showed no invasive fungi and no inflammatory infiltrates or damage (Fig. 1g). We detected very low numbers of *ece1 Δ /* cells in only 3/20 mice

¹Mucosal & Salivary Biology Division, Dental Institute, King's College London SE1 1UL, UK. ²Department of Microbial Pathogenicity Mechanisms, Hans Knöll Institute, D-07745 Jena, Germany.

³Research Center Borstel, Division of Biophysics, D-23845 Borstel, Germany. ⁴Deutsches Elektronen-Synchrotron DESY, D-22607 Hamburg, Germany. ⁵Department of Molecular & Biomedical Sciences, University of Maine, Orono, Maine 04469, USA. ⁶Wolfson CARD, King's College London, Guy's Campus, London SE1 1UL, UK. ⁷Research Group Microbial Immunology, Hans Knöll Institute, D-07745 Jena, Germany. ⁸Centre for Ultrastructural Imaging, King's College London, London SE1 1UL, UK. ⁹Department of Life Sciences, Imperial College London, London SW7 2AZ, UK. ¹⁰Septomics Research Center, Hans-Knöll Institute and Friedrich Schiller University, D-07745 Jena, Germany. ¹¹Department of Molecular and Applied Microbiology, Hans Knöll Institute, D-07745 Jena, Germany. ¹²Institute for Medical Microbiology, University Medical Center Göttingen, D-37075 Göttingen, Germany. ¹³Friedrich Schiller University, D-07737 Jena, Germany.

¹⁴Integrated Research and Treatment Center, Center for Sepsis Control and Care, D-07747 Jena, Germany. [†]Present addresses: Aberdeen Fungal Group, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen AB25 2ZD, UK (D.W.); NIHR Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust, London SE1 1UL, UK (M.R.); ERI Biotechmed & Microbiology and Ecology Department, University of Valencia, Valencia 46100, Spain (C.M.).

*These authors contributed equally to this work.

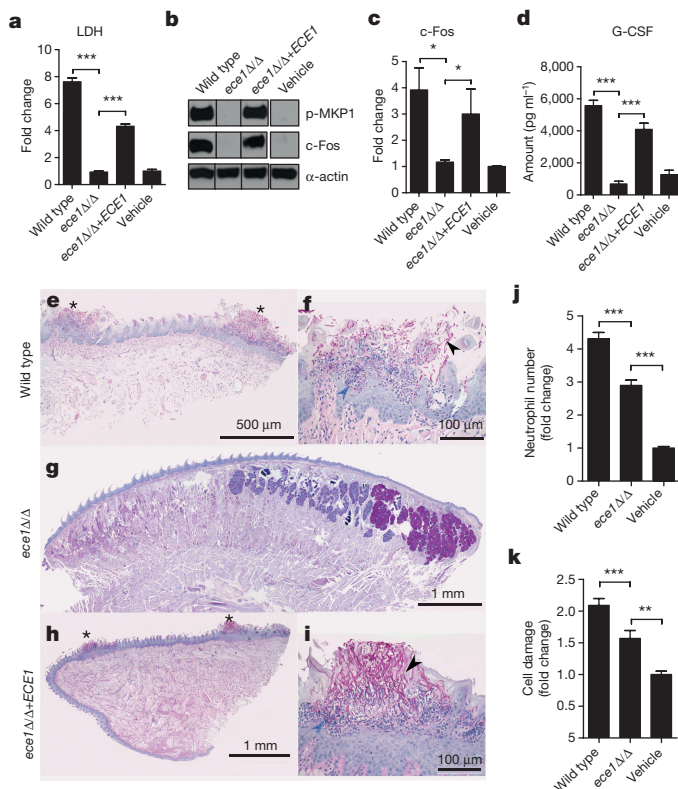


Figure 1 | *ECE1* is required for epithelial activation and *C. albicans* infection. TR146 cells were infected with the indicated *C. albicans* strains.

a, LDH release 24 h post-infection (p.i.) (MOI = 0.1). **b**, Induction of p-MKP-1 and c-Fos at 2 h p.i. (MOI = 10). **c**, c-Fos DNA binding at 3 h p.i. (MOI = 10). **d**, G-CSF production at 24 h p.i. (MOI = 0.01). **e–i**, PAS-stained tongues from mice subjected to OPC 2 days p.i. Whole-mount ($\times 25$) (**e, g, h**) and high-magnification (**f, i**) ($\times 200$) views of PAS-stained tongues of mice infected with *C. albicans* wild type (**e, f**), *ece1* Δ/Δ (**g**) and *ece1* Δ/Δ + *ECE1* (**h, i**). Invading hyphae (black arrowhead) and inflammatory cells (blue arrowhead) are indicated. **j**, Quantification of neutrophils in zebrafish swimbladder following infection with wild-type *C. albicans* (number of fish (n) = 47), *ece1* Δ/Δ (n = 53) or PBS (n = 40). **k**, Quantification of damaged cells in zebrafish swimbladder after infection with *C. albicans* wild-type (n = 73) or *ece1* Δ/Δ strains (n = 59) or vehicle (n = 63). Data are representative (**b, e–i**) or the mean (**a, c, d, j, k**) of three biological replicates. Error bars \pm s.e.m. Data were analysed by one-way ANOVA (**a, d**), paired *t*-test (**c**) or Kruskal–Wallis test (**j, k**). * P < 0.05, ** P < 0.01, *** P < 0.001. For gel image, see Supplementary Fig. 1.

(Extended Data Fig. 2a), which showed no evidence of local epithelial damage (data not shown). Quantification of histology sections indicated that the percentage of epithelial surface infected was significantly greater with the wild-type and *ECE1* re-integrand strains (Extended Data Fig. 2b). In a zebrafish swimbladder model of mucosal infection^{18,19}, neutrophil recruitment and tissue damage were both significantly lower following *ece1* Δ/Δ infection as compared with the wild-type strain (Fig. 1j, k and Extended Data Fig. 2c, d). Therefore, *C. albicans* Ece1p is critical for mucosal pathogenesis and is an innate immune activator *in vivo*.

Ecelp encodes a cytolytic peptide toxin

Ecelp is an *in vitro* substrate for Kex2p, a Golgi-located protease that cleaves proteins after lysine–arginine (KR) motifs²⁰. Ecelp contains seven KR-processing sites, suggesting it has the potential to produce eight secreted peptides from *C. albicans*²⁰ (Extended Data Fig. 3a, b). Liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis confirmed that recombinant Kex2p (rKex2p) processes recombinant Ecelp (rEcelp), and that all eight peptides generated terminated in KR (and fragments thereof, showing that less efficient processing

occurs also after a single K or R) (Supplementary Information). The importance of Kex2p-mediated Ecelp processing was demonstrated using a *kex2* Δ/Δ null strain²¹, which was unable to damage oral epithelia or induce p-MKP1/c-Fos-mediated danger responses or cytokine secretion (Extended Data Table 1). To determine which Ecelp peptide(s) were responsible for epithelial activation and damage, oral epithelial cells were incubated with peptides Ece1–I–VIII (1.5–70 μ M). Only Ece1–III_{62–93} induced p-MKP1, c-Fos, cytokines and damage (Fig. 2a–c, Extended Data Fig. 3c–e). Notably, low Ece1–III_{62–93} concentrations (1.5–15 μ M) were sufficient to induce c-Fos DNA binding (Fig. 2d), G-CSF and GM-CSF (Fig. 2c and Extended Data Fig. 3c), whereas high Ece1–III_{62–93} concentrations (70 μ M) were required to induce damage (Fig. 2e) and the damage-associated cytokines IL-1 α and IL-6, respectively (Extended Data Fig. 3d, e). Ece1–III_{62–93} could also directly lyse multiple human epithelial cell types and induce haemolysis of red blood cells, a classical test for cytotoxin activity (not shown). Neither the N-terminal hydrophobic region (Ece1–III_{62–85}) nor the C-terminal hydrophilic region (Ece1–III_{86–93}) induced p-MKP1, c-Fos, cytokines or damage of epithelial cells, either individually or in combination (Extended Data Fig. 3f–h), demonstrating that the peptide containing both regions is required for activity. Therefore, Ece1–III_{62–93} is the active region of Ecelp, acting as an epithelial immune activator and a cytolytic agent.

To confirm that Ece1–III_{62–93} drives epithelial activation and fungal pathogenicity, we generated a *C. albicans* strain lacking only the Ece1–III_{62–93} region (*ece1* Δ/Δ + *ECE1*_{184–279}). LC–MS/MS analysis showed that the modified protein in this strain is stable, secreted and processed into each of the predicted peptide fragments, with the exception of the deleted peptide toxin (Supplementary Information). Like *ece1* Δ/Δ , *ece1* Δ/Δ + *ECE1*_{184–279} efficiently formed invasive hyphae (not shown). However, *ece1* Δ/Δ + *ECE1*_{184–279} was unable to induce p-MKP1, c-Fos DNA binding, cytokines, or damage epithelia (Extended Data Fig. 3i–l). In murine OPC, unlike the *ece1* Δ/Δ + *ECE1* complemented strain, *ece1* Δ/Δ + *ECE1*_{184–279}-infected mice demonstrated absent (n = 4/10) or low (n = 6/10) fungal burdens, with no evidence of inflammatory infiltrates or local epithelial damage (Fig. 2f–h and Extended Data Fig. 4a, b). Likewise, *ece1* Δ/Δ + *ECE1*_{184–279} did not induce full damage in the zebrafish swimbladder model (Fig. 2i and Extended Data Fig. 4c). In contrast, injection of lytic doses of Ece1–III_{62–93} into the swimbladder induced epithelial damage (Fig. 2j, k). Thus, Ece1–III_{62–93} is both necessary and sufficient for epithelial immune activation, damage and mucosal infection *in vivo*.

The amphipathic properties of Ece1–III_{62–93} (SIIGIIMGILGNIPQVIQIIMSIVKAFKGNKR) coupled with the α -helical structure of the N-terminal hydrophobic region (Extended Data Fig. 5a, b) indicated that this fungal peptide may act similarly to cationic antimicrobial peptides and peptide toxins such as melittin²² (honey bee), magainin 2 (ref. 23) (African clawed frog) and alamethicin²⁴ (*Trichoderma viride*). Cytolytic peptide toxins have not previously been found in human pathogenic fungi, but bacterial cytolytic toxins are known to induce lesions after binding to target cell membranes^{25,26}. To investigate the importance of lipid composition for Ece1–III_{62–93}-mediated cytotoxicity, we used Förster resonance energy transfer (FRET) and electrical impedance spectroscopy to analyse the interactions of Ece1–III_{62–93} with model membranes comprised of lipid bilayers of dioleoylphosphatidylcholine (DOPC) with or without cholesterol. Although Ece1–III_{62–93} was able to efficiently intercalate into DOPC membranes (Extended Data Fig. 5c), Ece1–III_{62–93} permeabilization was enhanced in the presence of cholesterol (Fig. 3a). Ece1–III_{62–93}-induced lesions were heterogeneous and transient (Extended Data Fig. 5d), indicating that the peptide may damage target membranes through a ‘carpet-like’ mechanism²⁷. Patch-clamp analysis of epithelial cells demonstrated that lesion formation by Ece1–III_{62–93} is rapid and causes an inward current (Fig. 3b), associated with calcium influx (Fig. 3c). Similar phenomena occur with bacterial cytolytic toxins, which are known to trigger cell activation^{25,26,28}.

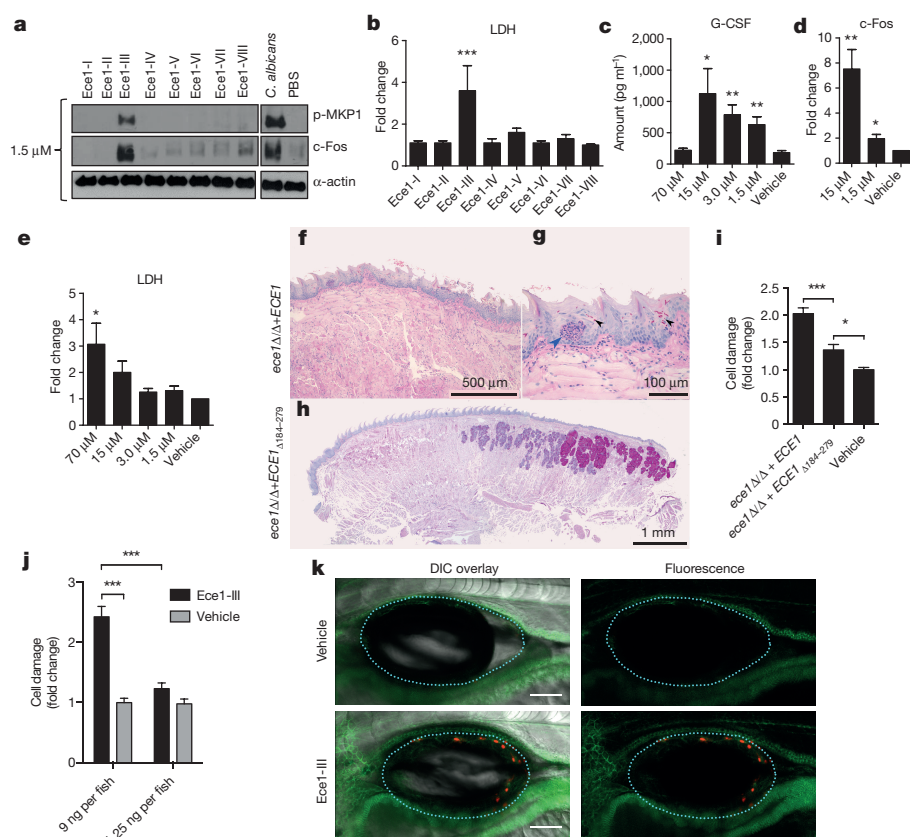


Figure 2 | Ece1-III₆₂₋₉₃ is the active region of Ece1p and is required for TR146 cell activation and mucosal *C. albicans* infection. **a**, Induction of p-MKP-1 and c-Fos 2 h post-stimulation (p.s.) with Ece1 peptides at 1.5 μM. **b**, LDH release 24 h p.s. with 70 μM Ece1 peptides. **c**, Induction of G-CSF 24 h p.s. with Ece1-III₆₂₋₉₃. **d**, c-Fos DNA binding induction 3 h p.s. with sub-lytic concentrations of Ece1-III₆₂₋₉₃. **e**, LDH release 24 h p.s. with Ece1-III₆₂₋₉₃. **f-h**, PAS-stained tongue sections from mice subjected to OPC, 2 days p.i. with *C. albicans* *ece1Δ/Δ* + *ECE1* (×25 and ×200) (f, g) or *ece1Δ/Δ* + *ECE1*_{Δ184-279} (h). Invading hyphae (black arrowheads) and infiltrating inflammatory cells (blue arrowhead) are shown. **i**, Damaged cells in a zebrafish swimbladder 24 h p.i. with *C. albicans* *ece1Δ/Δ* + *ECE1* (number of fish (n) = 44), *ece1Δ/Δ* + *ECE1*_{Δ184-279} (n = 58) or vehicle (n = 58). **j**, Damaged cells in zebrafish swimbladders after stimulation with 9 ng (n = 51) or 1.25 ng (n = 56) Ece1-III₆₂₋₉₃, or vehicle (40% DMSO, n = 54 and 5% DMSO, n = 55). **k**, Co-localization of adherens junctions (α-catenin-citrine) with Ece1-III₆₂₋₉₃-damaged cells (Sytox Orange-positive cells) in a zebrafish swimbladder. Scale bar, 100 μm. Data are representative (a, f-h, k) or mean (b-e, i, j) of three biological replicates (a-d) or ten mice or fish (f-h, k). Error bars show ± s.e.m. Data were analysed by one-way ANOVA (b, c, e) paired *t*-test (d) or Kruskal-Wallis (i, j). **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (compared with vehicle control unless otherwise indicated). For gel image, see Supplementary Fig. 1.

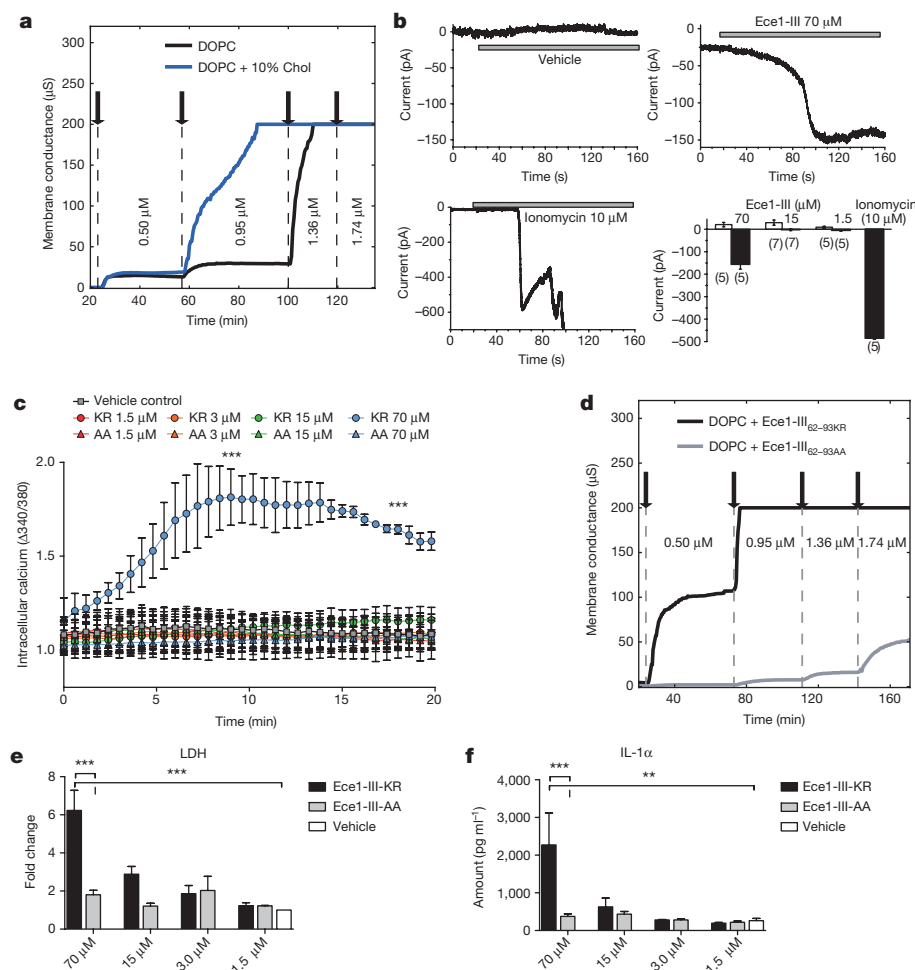


Figure 3 | Ece1-III₆₂₋₉₃ functions as a cytolytic peptide toxin. **a**, Kinetic changes in conductance of tethered lipid membranes after exposure to different concentrations of Ece1-III₆₂₋₉₃. **b**, Evoked inward current at a membrane potential of -60 mV in TR146 cells post-addition of Ece1-III₆₂₋₉₃ or ionomycin (positive control); individual (representative) and cumulative changes (bar chart indicates number of cells analysed below each bar) shown. **c**, Intracellular calcium level kinetics in TR146 cells post-stimulation (p.s.) with Ece1-III₆₂₋₉₃ wild type (Ece1-III_{62-93KR}) or Ece1-III₆₂₋₉₃ AA C-terminal substitution (Ece1-III_{62-93AA}). **d**, Kinetic changes in conductance of tethered DOPC membranes after exposure to different concentrations of Ece1-III_{62-93KR} and Ece1-III_{62-93AA}. **e**, **f**, LDH release (e) and secretion of IL-1α (f) from TR146 cells 24 h p.s. with Ece1-III_{62-93KR} or Ece1-III_{62-93AA}. Data shown are representative (a, d) or mean (b, c, e, f) of three biological replicates. Error bars show ± s.e.m. Data were analysed by one-way ANOVA (c, e and f). ***P* < 0.01, ****P* < 0.001.

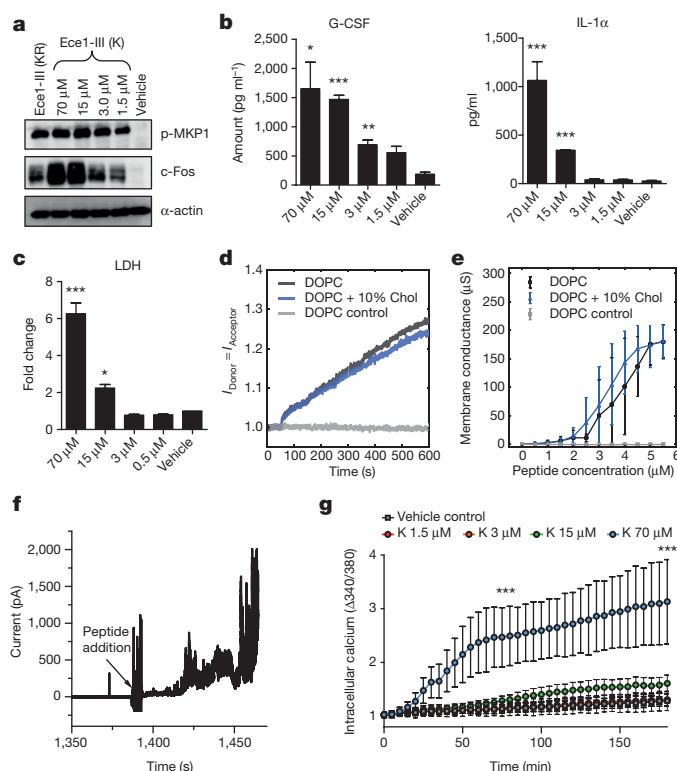


Figure 4 | Ece1-III_{62-92K} functions as a cytolytic peptide toxin that activates and damages epithelial cells. **a–c**, Induction of p-MKP-1 and c-Fos 2 h post-stimulation (p.s.) (**a**), secretion of G-CSF and IL-1- α 24 h p.s. (**b**), and LDH release 24 h p.s. of TR146 cells with Ece1-III_{62-92K} (**c**). **d**, Förster resonance energy transfer (FRET) showing intercalation of Ece1-III_{62-92K} (10 μ M) into lipid liposomes. **e**, Average peptide concentration-dependent changes in conductance of tethered lipid membranes. **f**, Ece1-III_{62-92K} (4 μ M) induced permeabilization of planar lipid membranes showing heterogeneous and transient lesions leading to membrane rupture. **g**, Intracellular calcium level kinetics in TR146 cells p.s. with Ece1-III_{62-92K}. Data shown are representative (**a**, **d**, **f**) or mean (**b**, **c**, **e**, **g**) of three biological replicates. Error bars show \pm s.e.m. Data are analysed by one-way ANOVA (**b**, **c**). * P < 0.05, ** P < 0.01, *** P < 0.001 (compared with vehicle control). For gel image, see Supplementary Fig. 1.

We postulated that the positively-charged C-terminal KR residues of Ece1-III₆₂₋₉₃ might be critical for interacting with negatively-charged components of host membranes to promote lesion formation. Substitution of the KR motif to AA (alanine-alanine; Ece1-III_{62-93AA}) did not affect membrane intercalation (not shown) but significantly reduced the peptide's ability to permeabilize membranes, damage epithelial cells and induce calcium influx (Fig. 3c–e). Thus, the positive C terminus of Ece1-III₆₂₋₉₃ is critical for lesion formation and damage induction in epithelial membranes. Notably, Ece1-III_{62-93AA} still induced p-MKP1, c-Fos and the non-damage associated cytokine G-CSF (Extended Data Fig. 5e, f), but not the damage-associated cytokine IL-1 α (Fig. 3f), suggesting that Ece1-III_{62-93AA} can be recognized by epithelial immunity without damaging cells. The importance of this finding is that it means epithelial cells are not only responding to damage, but have evolved to specifically recognize the peptide.

Ece1-III_{62-92K} is a secreted cytolytic peptide toxin

To demonstrate that Ece1-III is generated during epithelial infection, we performed LC-MS/MS analysis on the secretome from wild-type *C. albicans* hyphae grown in the presence and absence of epithelial cells (Supplementary Information). Notably, Ece1-III was the only peptide detected in the presence of epithelial cells, indicating that the fungus secretes this toxin during mucosal infection. However, the predominant form of secreted Ece1-III terminated in a K residue

(SIIGIIMGILGNIPQVIQIIMSIVKAFKGNK; Ece1-III_{62-92K}) and not KR (SIIGIIMGILGNIPQVIQIIMSIVKAFKGNKR; Ece1-III_{62-93KR}) (Extended Data Table 3). In fungi, it is known that following Kex2p processing, many proteins are subsequently cleaved by Kex1p²⁹ (also in the Golgi), removing the C-terminal R. LC-MS/MS analysis on the hyphal secretome of a *kex1* Δ/Δ mutant demonstrated that the predominant peptide secreted terminates in KR (not K) (Supplementary Information). Therefore, Ece1p is also subject to ordered Kex2p/Kex1p processing. Accordingly, we confirmed that Ece1-III_{62-92K} functioned similarly to Ece1-III_{62-93KR} with respect to epithelial cell activation. Specifically, Ece1-III_{62-92K} is also α -helical (not shown) and induces c-Fos, p-MKP1, cytokines (IL-1 α , G-CSF), damage (LDH), membrane intercalation and permeabilization, and calcium influx (Fig. 4a–g). Thus, the dominant peptide secreted from *C. albicans* hyphae during mucosal infection is Ece1-III_{62-92K}, which acts as a cytolytic peptide toxin that activates epithelial cells.

Based on these data, we propose a model of *C. albicans* mucosal infection whereby invasive hyphae secrete Ece1-III_{62-92K} into a membrane-bound 'invasion pocket'^{30,31}, facilitating peptide accumulation (Extended Data Fig. 6). During early stages of infection, sub-lytic concentrations of Ece1-III_{62-92K} induce epithelial immunity by activating the 'danger response' pathway (p-MKP1/c-Fos), alerting the host to the transition from colonizing yeast to invasive, toxin-producing hyphae. As infection progresses, Ece1-III_{62-92K} levels accumulate and elicit direct tissue damage. Mechanistically, we propose that the asymmetric distribution of charge along the α -helix of Ece1-III_{62-92K} facilitates correct peptide orientation relative to the host membrane, enabling intercalation, permeabilization and calcium influx. Our data identifies *C. albicans* Ece1-III_{62-92K} as the first cytolytic peptide toxin in a human fungal pathogen and reveals the molecular mechanisms of epithelial damage and host recognition of this clinically important fungus. We propose the name 'Candidalysin' for this newly discovered fungal toxin.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 July 2015; accepted 26 February 2016.

Published online 30 March 2016.

- Brown, G. D. et al. Hidden killers: human fungal infections. *Sci. Transl. Med.* **4**, 165rv113 (2012).
- Jacobsen, I. D. et al. *Candida albicans* dimorphism as a therapeutic target. *Expert Rev. Anti Infect. Ther.* **10**, 85–93 (2012).
- Moyes, D. L. et al. A biphasic innate immune MAPK response discriminates between the yeast and hyphal forms of *Candida albicans* in epithelial cells. *Cell Host Microbe* **8**, 225–235 (2010).
- Moyes, D. L. et al. *Candida albicans* yeast and hyphae are discriminated by MAPK signaling in vaginal epithelial cells. *PLoS ONE* **6**, e26580 (2011).
- Murciano, C. et al. *Candida albicans* cell wall glycosylation may be indirectly required for activation of epithelial cell proinflammatory responses. *Infect. Immun.* **79**, 4902–4911 (2011).
- Moyes, D. L. et al. Activation of MAPK/c-Fos induced responses in oral epithelial cells is specific to *Candida albicans* and *Candida dubliniensis* hyphae. *Med. Microbiol. Immunol. (Berl.)* **201**, 93–101 (2012).
- Murciano, C. et al. Evaluation of the role of *Candida albicans* agglutinin-like sequence (Als) proteins in human oral epithelial cell interactions. *PLoS ONE* **7**, e33362 (2012).
- Moyes, D. L. & Naglik, J. R. Mucosal immunity and *Candida albicans* infection. *Clin. Dev. Immunol.* **2011**, 346307 (2011).
- Naglik, J. R. & Moyes, D. Epithelial cell innate response to *Candida albicans*. *Adv. Dent. Res.* **23**, 50–55 (2011).
- Naglik, J. R., Moyes, D. L., Wachtler, B. & Hube, B. *Candida albicans* interactions with epithelial cells and mucosal immunity. *Microbes Infect.* **13**, 963–976 (2011).
- Hebecker, B., Naglik, J. R., Hube, B. & Jacobsen, I. D. Pathogenicity mechanisms and host response during oral *Candida albicans* infections. *Expert Rev. Anti Infect. Ther.* **12**, 867–879 (2014).
- Naglik, J. R. *Candida* immunity. *New J. Science* **2014**, 390241 (2014).
- Naglik, J. R., Richardson, J. P. & Moyes, D. L. *Candida albicans* Pathogenicity and epithelial immunity. *PLoS Pathog.* **10**, e1004257 (2014).
- Moyes, D. L., Richardson, J. P. & Naglik, J. R. *Candida albicans*—epithelial interactions and pathogenicity mechanisms: scratching the surface. *Virulence* **6**, 338–346 (2015).

15. Birse, C. E., Irwin, M. Y., Fonzi, W. A. & Sypherd, P. S. Cloning and characterization of *ECE1*, a gene expressed in association with cell elongation of the dimorphic pathogen *Candida albicans*. *Infect. Immun.* **61**, 3648–3655 (1993).
16. Röhm, M. *et al.* A family of secreted pathogenesis-related proteins in *Candida albicans*. *Mol. Microbiol.* **87**, 132–151 (2013).
17. Kamaï, Y., Kubota, M., Hosokawa, T., Fukuoka, T. & Filler, S. G. New model of oropharyngeal candidiasis in mice. *Antimicrob. Agents Chemother.* **45**, 3195–3197 (2001).
18. Brothers, K. M. *et al.* NADPH oxidase-driven phagocyte recruitment controls *Candida albicans* filamentous growth and prevents mortality. *PLoS Pathog.* **9**, e1003634 (2013).
19. Gratacap, R. L., Rawls, J. F. & Wheeler, R. T. Mucosal candidiasis elicits NF- κ B activation, proinflammatory gene expression and localized neutrophilia in zebrafish. *Dis. Model. Mech.* **6**, 1260–1270 (2013).
20. Bader, O., Krauke, Y. & Hube, B. Processing of predicted substrates of fungal Kex2 proteinases from *Candida albicans*, *C. glabrata*, *Saccharomyces cerevisiae* and *Pichia pastoris*. *BMC Microbiol.* **8**, 116 (2008).
21. Newport, G. & Agabian, N. *KEX2* influences *Candida albicans* proteinase secretion and hyphal formation. *J. Biol. Chem.* **272**, 28954–28961 (1997).
22. Liu, P., Huang, X., Zhou, R. & Berne, B. J. Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature* **437**, 159–162 (2005).
23. Bechinger, B. & Salnikow, E. S. The membrane interactions of antimicrobial peptides revealed by solid-state NMR spectroscopy. *Chem. Phys. Lipids* **165**, 282–301 (2012).
24. Pieta, P., Mirza, J. & Lipkowski, J. Direct visualization of the alamethicin pore formed in a planar phospholipid matrix. *Proc. Natl Acad. Sci. USA* **109**, 21223–21227 (2012).
25. Bischofberger, M., Iacovache, I. & van der Goot, F. G. Pathogenic pore-forming proteins: function and host response. *Cell Host Microbe* **12**, 266–275 (2012).
26. Los, F. C., Randis, T. M., Aroian, R. V. & Ratner, A. J. Role of pore-forming toxins in bacterial infectious diseases. *Microbiol. Mol. Biol. Rev.* **77**, 173–207 (2013).
27. Oren, Z. & Shai, Y. Selective lysis of bacteria but not mammalian cells by diastereomers of melittin: structure-function study. *Biochemistry* **36**, 1826–1835 (1997).
28. Walev, I. *et al.* Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-O. *Proc. Natl Acad. Sci. USA* **98**, 3185–3190 (2001).
29. Schmitt, M. J. & Breinig, F. Yeast viral killer toxins: lethality and self-protection. *Nature Rev. Microbiol.* **4**, 212–221 (2006).
30. Zakikhany, K. *et al.* *In vivo* transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination. *Cell. Microbiol.* **9**, 2938–2954 (2007).
31. Wächter, B. *et al.* *Candida albicans*-epithelial interactions: dissecting the roles of active penetration, induced endocytosis and host factors on the infection process. *PLoS ONE* **7**, e36952 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Gaffen, B. Klein, C. Hertweck, A. Tucker, J. Green and S. Challacombe for comments on the manuscript. For experimental assistance, we thank S. Bevan and D. Andersson (calcium assays), D. Nayar (histology), D. Rahman and M. Mistry (murine model), M. Nilan (zebrafish model), S. Groth (FRET spectroscopy), N. Gebauer (Impedance experiments), D. Schulz (*kex1* Δ/Δ strain) and our colleagues for supplying fungal mutant strains. This work was supported by grants from the Medical Research Council (MR/J008303/1, MR/M011372/1), Biotechnology & Biological Sciences Research Council (BB/J015261/1), FP7-PEOPLE-2013-Initial Training Network (606786) to J.R.N.; Wellcome Trust Strategic Award for Medical Mycology and Fungal Immunology (097377/Z/11/Z) to J.R.N. and D.W.; Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (102549/Z/13/Z) to D.W.; Deutsche Forschungsgemeinschaft CRC/TR124 FungiNet Project C1 and Z2, Deutsche Forschungsgemeinschaft SPP 1580 (Hu 528/17-1) and CSCC, German Federal Ministry of Education and Health (BMBF) 01EO1002 to B.Hu.; Cluster of Excellence 'Inflammation at interfaces' and Deutsche Forschungsgemeinschaft SPP 1580 project GU 568/5-1 to T.G.; National Institutes of Health (R15AI094406) and the Burroughs Wellcome Fund to R.T.W.

Author Contributions D.L.M., J.P.R., S.X.T., M.R., C.M., M.B., S.I.I. and N.K. performed signalling, transcription factor, calcium and cytokine assays, and murine work; D.W., S.H., S.M., T.M.F., B.He., L.K. A.H., O.B. and O.Ku. created fungal strains and performed fluorescent microscopy, adhesion, invasion, gene expression and damage assays; R.L.G. and R.T.W. performed zebrafish experiments; J.W. and T.G. performed biophysical analysis with artificial membranes; J.R. performed whole patch clamp analysis; G.V. performed electron microscopy; S.T. performed histological analysis; S.M., T.L., T.K. and O.Kn. performed LC-MS analyses; J.R.N., B.Hu., D.L.M., J.P.R. and D.W. wrote the paper; J.R.N., B.Hu. and E.C. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.Hu. (bernhard.hube@leibniz-hki.de).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines, reagents and *Candida* strains. Experiments were carried out using the TR146 buccal epithelial squamous cell carcinoma line³² obtained from the European Collection of Authenticated Cell Cultures (ECACC) and grown in Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Cells were routinely tested for mycoplasma contamination using mycoplasma-specific primers and were found to be negative. Prior to stimulation, confluent TR146 cells were serum-starved overnight, and all experiments were carried out in serum-free DMEM. *C. albicans* wild-type strains included the autotrophic strain BWP17 + Clp30 (ref. 33) and the parental strain SC5314 (ref. 34). Other *C. albicans* strains used and their sources are listed in Extended Data Tables 1 and 2. *C. albicans* cultures were grown in YPD medium (1% yeast extract, 2% peptone, 2% dextrose) at 30 °C overnight. Cultures were washed in sterile PBS and adjusted to the required cell density. Antibodies to phospho-MKP1 and c-Fos were from Cell Signalling Technologies (New England Biolabs UK), mouse anti-human α -actin was from Millipore (UK), and goat anti-mouse and anti-rabbit horseradish peroxidase (HRP)-conjugated antibodies were from Jackson Immunologicals (Strattech Scientific, UK). Ece1p peptides were synthesized commercially (Proteogenix (France) or Peptide Synthetics (UK)).

Generation of *C. albicans* ECE1 mutant strains. ECE1 deletion was performed as previously described³⁵. Deletion cassettes were generated by PCR³⁶. Primers ECE1-FG and ECE1-RG were used to amplify pFA-HIS1 and pFA-ARG4-based markers. *C. albicans* BWP17 (ref. 37), was sequentially transformed³⁸ with the ECE1-HIS1 and ECE1-ARG4 deletion cassettes and then transformed with Clp10 (ref. 39), yielding the *ece1* Δ/Δ deletion strain. For complementation, the ECE1 gene plus upstream and downstream intergenic regions were amplified with primers ECE1-RecF3k and ECE1-RecR and cloned into plasmid Clp10 at MluI and SalI sites. This plasmid was transformed into the uridine auxotrophic *ece1* Δ/Δ strain, yielding the *ece1* Δ/Δ + ECE1 complemented strain. For generation of the *ece1* Δ/Δ + ECE1 $\Delta_{184-279}$ strain, the Clp10-ECE1 was amplified with primers Pep3-F1 and Pep3-R1, digested with ClaI and re-ligated, yielding the Clp10 + ECE1 $\Delta_{184-279}$ plasmid. This plasmid was transformed into the uridine auxotrophic *ece1* Δ/Δ strain, yielding the *ece1* Δ/Δ + ECE1 $\Delta_{184-279}$ strain. All integrations were confirmed by PCR/sequencing and at least two independent isogenic transformants were created to confirm results. KEX1 deletion was performed exactly as the ECE1 deletion but using primers KEX1-FG and KEX1-RG for creating the deletion cassette. Fluorescent strains of *ece1* Δ/Δ and BWP17 were constructed as previously described⁴⁰. Briefly, the *ece1* Δ/Δ and BWP17 strains were transformed with the pENO1-dTom-NATr plasmid. Primers used to clone and construct the ECE1 genes and intragenic regions are listed in Extended Data Table 4. Strains are listed in Extended Data Table 2.

Construction of *C. albicans* ECE1 promoter-GFP strain. ECE1 promoter (primers 5'ECE1prom-NarI / 3'ECE1prom-XhoI) and terminator (5'ECE1term-SacII / 5'ECE1term-SacI) were amplified and cloned into pADH1-GFP. Resulting pSK-pECE1-GFP was verified by sequencing. *C. albicans* SC5314 was transformed with the pECE1-GFP transformation cassette³⁸. Resistance to nourseothricin was used as selective marker and correct integration of GFP into the ECE1 locus was verified by PCR. Primers for cloning and validation are listed in Extended Data Table 4. Strains are listed in Extended Data Table 2.

RNA isolation and real-time PCR analysis. *C. albicans* cells grown on TR146 epithelial cells were collected into RNA pure (PeqLab), centrifuged and the pellet resuspended in 400 μ l AE buffer (50 mM Na-acetate pH 5.3, 10 mM EDTA, 1% SDS). Samples were vortexed (30 s), and an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added and incubated for 5 min (65 °C) before subjected to 2 \times freeze-thawing. Lysates were clarified by centrifugation and the RNA precipitated with isopropyl alcohol/0.3 M sodium acetate by incubating for 1 h at -20 °C. Precipitated pellets were washed (2 \times 1 ml 70% ice-cold ethanol), resuspended in DEPC-treated water and stored at -80 °C. RNA integrity and concentration was confirmed using a Bioanalyzer (Agilent). RNA (500 ng) was treated with DNase (Epicentre) and cDNA synthesized using Reverse Transcriptase Superscript III (Invitrogen). cDNA samples were used for qPCR with EVAgreen mix (Bio&Sell). Primers (ACT1-F and ACT1-R for actin, ECE1-F and ECE1-R for ECE1 Extended Data Table 4) were used at a final concentration of 500 nM. qPCR amplifications were performed using a Biorad CFX96 thermocycler. Data was evaluated using Bio-Rad CFX Manager 3.1 (Bio-Rad) with ACT1 as the reference gene and t_0 as the control sample.

Western blotting. TR146 cells were lysed using a modified RIPA lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1%

sodium deoxycholate, 0.1% SDS) containing protease (Sigma-Aldrich) and phosphatase (Perbio Science) inhibitors⁴¹, left on ice (30 min) and then clarified (10 min) in a refrigerated microfuge. Lysate total protein content was determined using the BCA protein quantitation kit (Perbio Science). 20 μ g of total protein was separated on 12% SDS-PAGE gels before transfer to nitrocellulose membranes (GE Healthcare). After probing with primary (1:1,000) and secondary (1:10,000) antibodies, membranes were developed using Immobilon chemiluminescent substrate (Millipore) and exposed to X-ray film (Fuji film). Human α -actin was used as a loading control.

Transcription factor DNA binding assay. DNA binding activity of transcription factors was assessed using the TransAM transcription factor ELISA system (Active Motif) as previously described^{41,42}. Serum-starved TR146 epithelial cells were treated for 3 h before being differentially lysed to recover nuclear proteins using a nuclear protein extraction kit (Active Motif) according to the manufacturer's protocol. Protein concentration was determined (BCA protein quantitation kit (Perbio Science)) and 5 μ g of nuclear extract was assayed in the TransAM system according to the manufacturer's protocol. Data was expressed as fold-change in $A_{450\text{ nm}}$ relative to resting cells.

Cytokine determination. Cytokine levels in cell culture supernatants were determined using the Performance magnetic Fluorokine MAP cytokine multiplex kit (Bio-technique) and a Bioplex 200 machine. The data were analysed using Bioplex Manager 6.1 software to determine analyte concentrations.

Cell damage assay. Following incubation, culture supernatant was collected and assayed for lactate dehydrogenase (LDH) activity using the Cytox 96 Non-Radioactive Cytotoxicity Assay kit (Promega) according to the manufacturer's instructions. Recombinant porcine LDH (Sigma-Aldrich) was used to generate a standard curve.

Epithelial adhesion assay. Quantification of *C. albicans* adherence to TR146 epithelial cells was performed as described previously⁴³. Briefly, TR146 cells were grown to confluence on glass coverslips for 48 h in tissue culture plates in DMEM medium. *C. albicans* yeast cells (2×10^5) were added into 1 ml serum-free DMEM, incubated for 60 min (37 °C/5% CO₂) and non-adherent *C. albicans* cells removed by aspiration. Following washing (3 \times 1 ml PBS), cells were fixed with 4% paraformaldehyde (Roth) and adherent *C. albicans* cells stained with Calcofluor White and quantified using fluorescence microscopy. The number of adherent cells was determined by counting 100 high-magnification fields of 200 μ m \times 200 μ m size. Exact total cell numbers were calculated based on the quantified areas and the total size of the cover slip.

Epithelial invasion assay. *C. albicans* invasion of epithelial cells was determined as described previously⁴³. Briefly, TR146 epithelial cells were grown to confluence on glass coverslips for 48 h and then infected with *C. albicans* yeast cells (1×10^5), for 3 h in a humidified incubator (37 °C/5% CO₂). Following washing (3 \times PBS), the cells were fixed with 4% paraformaldehyde. All surface adherent fungal cells were stained for 1 h with a rabbit anti-*Candida* antibody and subsequently with a goat anti-rabbit-Alexa Fluor 488 antibody. After rinsing with PBS, epithelial cells were permeabilized (0.1% Triton X-100 in PBS for 15 min) and fungal cells (invading and non-invading) were stained with Calcofluor White. Following rinsing with water, coverslips were visualized using fluorescence microscopy. The percentage of invading *C. albicans* cells was determined by dividing the number of (partially) internalized cells by the total number of adherent cells. At least 100 fungal cells were counted on each coverslip.

Imaging of *C. albicans* growth and invasion of epithelial cells. TR146 cells (10^5 per ml) seeded on glass coverslips in DMEM/10% FBS were infected with *C. albicans* (2.5×10^4 cfu per ml) in DMEM and incubated for 6 h (37 °C/5% CO₂). Cells were washed with PBS, fixed overnight (4 °C in 4% paraformaldehyde) and stained with Concanavalin A-Alexa Fluor 647 in PBS (10 μ g ml⁻¹) for 45 min at room temperature in the dark with gentle shaking (70 r.p.m.) to stain the fungal cell wall. Epithelial cells were permeabilized with 0.1% Triton X-100 for 15 min at 37 °C in the dark, then washed and stained with 10 μ g ml⁻¹ Calcofluor White (0.1 M Tris-HCl pH 9.5) for 20 min at room temperature in the dark with gentle shaking. Cells were rinsed in water and mounted on slides with 6 μ l of ProLong Gold anti-fade reagent, before air drying for 2 h in the dark. Fluorescence microscopy was performed on a Zeiss Axio Observer Z1 microscope, and 5 phase images were taken per picture.

Scanning electron microscopy. For scanning electron microscopy (SEM) analysis, TR146 cells were grown to confluence on Transwell inserts (Greiner) and serum starved overnight in serum-free DMEM. After 5 h of *C. albicans* incubation on epithelial cells at an MOI of 0.01, cell media was removed and samples were fixed overnight at 4 °C with 2.5% (v/v) glutaraldehyde in 0.05 M HEPES buffer (pH 7.2) and post-fixed in 1% (w/v) osmium tetroxide for 1 h at room temperature. After washing, samples were dehydrated through a graded ethanol series before being critical point dried (Polaron E3000, Quorum Technologies). Dried samples were

mounted using carbon double side sticky discs (TAAB) on aluminium pins (TAAB) and gold coated in an Emitech K550X sputter coater (Quorum Technologies Ltd). Samples were examined and images recorded using a FEI Quanta 200 field emission scanning electron microscope operated at 3.5 kV in high vacuum mode.

Zebrafish swimbladder mucosal infection model. Zebrafish infections were performed in accordance with NIH guidelines under Institutional Animal Care and Use Committee (IACUC) protocol A2009-11-01 at the University of Maine. To determine sample size, a power calculation was done for all experiments based on two-tailed *t*-tests in order to detect a minimum effect size of 0.8, with an alpha error probability of 0.05 and a power (1 – beta error probability) of 0.95. This gave a minimum number of 42 fish for each group. The fish selected for the experiments were randomly assigned to the different groups by picking them from a pool without bias and the groups were injected in different orders. No blinding was used to read the results. Ten to twenty zebrafish per group per experiment were maintained at 33 °C in E3 + PTU and used as previously described⁴⁰. Briefly, 4 days post-fertilization (dpf) larvae were treated with 20 µg ml⁻¹ dexamethasone dissolved in 0.1% DMSO 1 h before infection and thereafter. For tissue damage and neutrophil recruitment, individual AB or *mpo:GFP* fish (respectively) were injected into the swimbladder with 4 nl of PBS with or without 25–40 *C. albicans* yeast cells of *ece1Δ/Δ-dTomato*, *ece1Δ/Δ + ECE1 + dTomato*, *ece1Δ/Δ + ECE1_{Δ184-279} + dTomato* or BWP17-*dTomato*. For tissue damage, 1 nl of Sytox green (0.05 mM in 1% DMSO) was injected at 20 h post-infection into the swimbladder and fish were imaged by confocal microscopy at 24 h post-infection. For neutrophil recruitment, fish were imaged at 24 h post-injection. For synthetic peptide damage, AB or α-catenin: citrine⁴⁴ fish were injected with 2 nl of peptide (9 ng or 1.25 ng per fish) or vehicle (40% DMSO or 5% DMSO) + SytoxGreen (0.05 mM in 1% DMSO) or SytoxOrange (0.5 mM in 10% DMSO) and the fish imaged by confocal microscopy 4 h later. Numbers of neutrophils and damaged cells observed were counted and tabulated for each fish.

Zebrafish swimbladder fluorescence microscopy. Live zebrafish imaging was carried out as previously described⁴⁰. Briefly, fish were anaesthetized in Tris-buffered Tricaine (200 µg ml⁻¹, Western Chemicals) and further immobilized in a solution of 0.4% low-melting-point agarose (LMA, Lonza) in E3 + Tricaine in a 96-well plate glass-bottom imaging dish (Greiner Bio-On). Confocal imaging was carried out using an Olympus IX-81 inverted microscope with an FV-1000 laser scanning confocal system (Olympus). Images were collected and processed using Fluoview (Olympus) and Photoshop (Adobe Systems). Panels are either a single slice for the differential interference contrast channel (DIC) with maximum projection overlays of fluorescence image channels (red-green), or maximum projection overlays of fluorescence channels. The number of slices for each maximum projection is specified in the legend of individual figures.

Murine oropharyngeal candidiasis model. Murine infections were performed under UK Home Office Project Licence PPL 70/7598 in dedicated animal facilities at King's College London. No statistical method was used to pre-determine sample size. No method of randomization was used to allocate animals to experimental groups. Mice in the same cage were part of the same treatment. The investigators were not blinded during outcome assessment. A previously described murine model of oropharyngeal candidiasis using female BALB/c mice⁴⁵ was modified to use for investigating early infection events. Briefly, mice were treated subcutaneously with 3 mg per mouse (in 200 µl PBS with 0.5% Tween 80) of cortisone acetate on days –1 and +1 post-infection. On day 0, mice were sedated for ~75 min with an intra-peritoneal injection of 110 mg per kg ketamine and 8 mg per kg xylazine, and a swab soaked in a 10⁷ cfu per ml of *C. albicans* yeast culture in sterile saline was placed sublingually for 75 min. After 2 days, mice were euthanized, the tongue excised and divided longitudinally in half. One half was weighed, homogenized and cultured to derive quantitative *Candida* counts. The other half was processed for histopathology and immunohistochemistry.

Immunohistochemistry of murine tissue. *C. albicans*-infected murine tongues were fixed in 10% (v/v) formal-saline before being embedded and processed in paraffin wax using standard protocols. For each tongue, 5-µm sections were prepared using a Leica RM2055 microtome and silane coated slides. Sections were dewaxed using xylene, before *C. albicans* and infiltrating inflammatory cells were visualized by staining using Periodic Acid-Schiff (PAS) stain and counterstaining with haematoxylin. Sections were then examined by light microscopy. Histological quantification of infection was undertaken by measuring the area of infected epithelium and expressed as a percentage relative to the entire epithelial area.

Whole cell patch clamp. TR146 epithelial cells were grown in 35-mm Petri dishes (Nunc) for 48 h before recordings at low cell density (10–30% confluence). Cells were superfused with a modified Krebs solution (120 mM NaCl, 3 mM KCl, 2.5 mM CaCl₂, 1.2 mM MgCl₂, 22.6 mM NaHCO₃, 11.1 mM glucose, 5 mM HEPES pH 7.4). Isolated cells were recorded at room temperature (21–23 °C) in whole cell mode using microelectrodes (5–7 MΩ) containing 90 mM potassium acetate, 20 mM KCl, 40 mM HEPES, 3 mM EGTA, 3 mM MgCl₂, 1 mM CaCl₂ (free Ca²⁺ 40 nM),

pH 7.4. Cells were voltage clamped at –60 mV using an Axopatch 200A amplifier (Axon Instruments) and current/voltage curves were generated by 1 s steps between –100 to +50 mV. Treatments were applied to the superfusate to produce the final required concentration, with vehicle controls similarly applied. Data was recorded using Clampex software (PCLamp 6, Axon Instrument) and analysed with Clampfit 10.

Calcium flux. TR146 cells were grown in a 96-well plate overnight until confluent. The medium was removed and 50 µl of a Fura-2 solution (5 µl Fura-2 (Life Technologies) (2.5 mM in 50% Pluronic F-127 (Life Technologies):50% DMSO), 5 µl probenecid (Sigma) in 5 ml saline solution (NaCl (140 mM), KCl (5 mM), MgCl₂ (1 mM), CaCl₂ (2 mM), glucose (10 mM) and HEPES (10 mM), adjusted to pH 7.4)) was added and the plate incubated for 1 h at 37 °C/5% CO₂. The Fura-2 solution was replaced with 50 µl saline solution and baseline fluorescence readings (excitation 340 nm/emission 520 nm) taken for 10 min using a FlexStation 3 (Molecular Devices). Ece1 peptides were added at different concentrations and readings immediately taken for up to 3 h. The data was analysed using Softmax Pro software to determine calcium present in the cell cytosol and expressed as the ratio between excitation and emission spectra.

Impedance spectroscopy of tethered bilayer lipid membranes (tBLMs). tBLMs with 10% tethering lipids and 90% spacer lipids (T10 slides) were formed using the solvent exchange technique^{46,47} according to the manufacturer's instructions (SDx Tethered Membranes Pty Ltd, Sydney, Australia). Briefly, 8 µl of 3 mM lipid solutions in ethanol were added, incubated for 2 min and then 93.4 µl buffer (100 mM KCl, 5 mM HEPES, pH 7.0) was added. After rinsing 3 × with 100 µl buffer the conductance and capacitance of the membranes were measured for 20 min before injection of Ece1 peptides at different concentrations. All experiments were performed at room temperature. Signals were measured using the tethaPod (SDx Tethered Membranes Pty, Sydney, Australia).

FRET intercalation experiments. Intercalation of Ece1 peptides into phospholipid liposomes was determined by FRET spectroscopy applied as a probe-dilution assay⁴⁸. Phospholipids mixed with each 1% (mol/mol) of the donor dye NBD-phosphatidylethanolamine (NBD-PE) and of the acceptor dye rhodamine-PE, were dissolved in chloroform, dried, solubilized in 1 ml buffer (100 mM KCl, 5 mM HEPES, pH 7.0) by vortexing, sonicated with a titan tip (30 W, Branson sonifier, cell disruptor B15), and subjected to three cycles of heating to 60 °C and cooling down to 4 °C, each for 30 min. Lipid samples were stored at 4 °C for at least 12 h before use. Ece1 peptide was added to liposomes and intercalation was monitored as the increase of the quotient between the donor fluorescence intensity *I*_D at 531 nm and the acceptor intensity *I*_A at 593 nm (FRET signal) independent of time.

Circular Dichroism spectroscopy. CD measurements were performed using a Jasco J-720 spectropolarimeter (Japan Spectroscopic Co., Japan), calibrated as described previously⁴⁹. CD spectra represent the average of four scans obtained by collecting data at 1 nm intervals with a bandwidth of 2 nm. The measurements were performed in 100 mM KCl, 5 mM HEPES, pH 7.0 at 25 °C and 40 °C in a 1.0 mm quartz cuvette. The Ece1-III concentration was 15 µM.

Planar lipid bilayers. Planar lipid bilayers were prepared using the Montal-Mueller technique⁵⁰ as described previously⁵¹. All measurements were performed in 5 mM HEPES, 100 mM KCl, pH 7.0 (specific electrical conductivity 17.2 mS per cm) at 37 °C.

Hyphal secretome preparation for LC-MS/MS analysis. *Candida* strains were cultured for 18 h in hyphae inducing conditions (YNB medium containing 2% sucrose, 75 mM MOPS buffer pH 7.2, 5 mM N-acetyl-D-glucosamine, 37 °C). Hyphal supernatants were collected by filtering through a 0.2 µm PES filter, and peptides were enriched by solid phase extraction (SPE) using first C4 and subsequently C18 columns on the C4 flowthrough. After drying in a vacuum centrifuge, samples were resolubilized in loading solution (0.2% formic acid in 71:27:2 ACN/H₂O/DMSO (v/v/v)) and filtered through a 10 kDa MWCO filter. The filtrate was transferred into HPLC vials and injected into the LC-MS/MS system. LC-MS/MS analysis was carried out on an Ultimate 3000 nano RSLC system coupled to a QExactive Plus mass spectrometer (ThermoFisher Scientific). Peptide separation was performed based on a direct injection setup without peptide trapping using an Accucore C4 column as stationary phase and a column oven temperature of 50 °C. The binary mobile phase consisting of A) 0.2% (v/v) formic acid in 95:5 H₂O/DMSO (v/v) and B) 0.2% (v/v) formic acid in 85:10:5 ACN/H₂O/DMSO (v/v/v) was applied for a 60 min gradient elution: 0–1.5 min at 60% B, 35–45 min at 96% B, 45.1–60 min at 60% B. The Nanospray Flex Ion Source (ThermoFisher Scientific) provided with a stainless steel emitter was used to generate positively charged ions at 2.2 kV spray voltage. Precursor ions were measured in full scan mode within a mass range of *m/z* 300–1600 at a resolution of 70k FWHM using a maximum injection time of 120 ms and an automatic gain control target of 1e6. For data-dependent acquisition, up to 10 most abundant precursor ions per scan cycle with an assigned charge state of *z* = 2–6 were selected in the quadrupole for

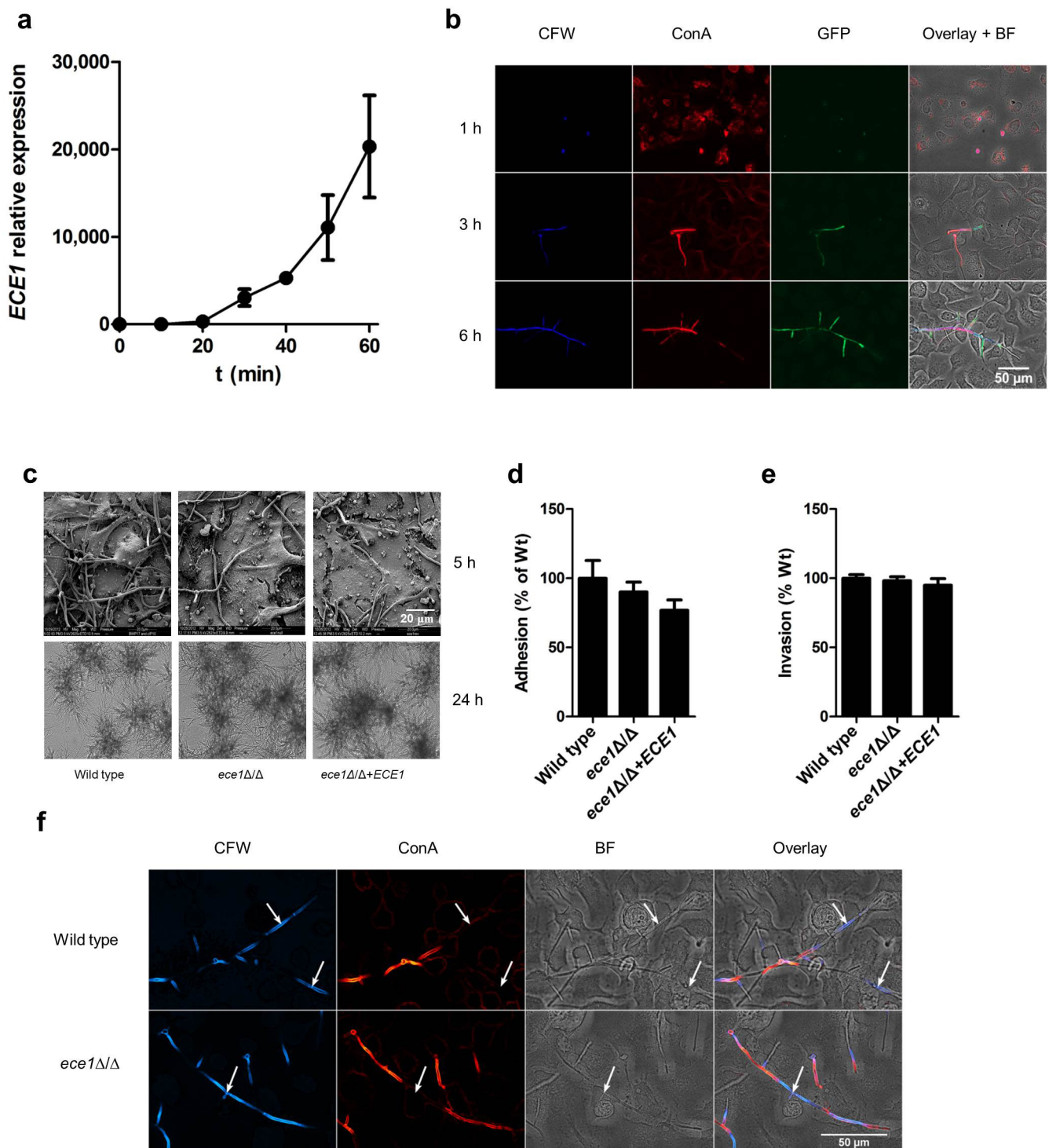
further fragmentation using an isolation width of m/z 2.0. Fragment ions were generated in the HCD cell at a normalized collision energy of 30 V using nitrogen gas. Dynamic exclusion of precursor ions was set to 20 s. Fragment ions were monitored at a resolution of 17.5k (FWHM) using a maximum injection time of 120 ms and an AGC target of 2e5.

Protein database search. Thermo raw files were processed by the Proteome Discoverer (PD) software v1.4.0.288 (Thermo). Tandem mass spectra were searched against the Candida Genome Database (http://www.candidagenome.org/download/sequence/C_albicans_SC5314/Assembly22/current/C_albicans_SC5314_A22_current_orf_trans_all.fasta.gz; status: 2015/05/03) using the Sequest HT search algorithm. Mass spectra were searched for both unspecific cleavages (no enzyme) and tryptic peptides with up to 4 missed cleavages. The precursor mass tolerance was set to 10 p.p.m. and the fragment mass tolerance to 0.02 Da. Target Decoy PSM Validator node and a reverse decoy database was used for (q value) validation of the peptide spectral matches (PSMs) using a strict target false discovery (FDR) rate of <1%. Furthermore, we used the score versus charge state function of the Sequest engine to filter out insignificant peptide hits (xcorr of 2.0 for $z=2$, 2.25 for $z=3$, 2.5 for $z=4$, 2.75 for $z=5$, 3.0 for $z=6$). At least two unique peptides per protein were required for positive protein hits.

Statistics. TransAM and patch clamp data were analysed using a paired *t*-test while cytokines, LDH and calcium influx data were analysed using one-way ANOVA with all compared groups passing an equal variance test. Murine *in vivo* data was analysed using the Mann–Whitney test. Zebrafish data was analysed using the Kruskal–Wallis test with Dunn's multiple comparison correction. In all cases, $P < 0.05$ was taken to be significant.

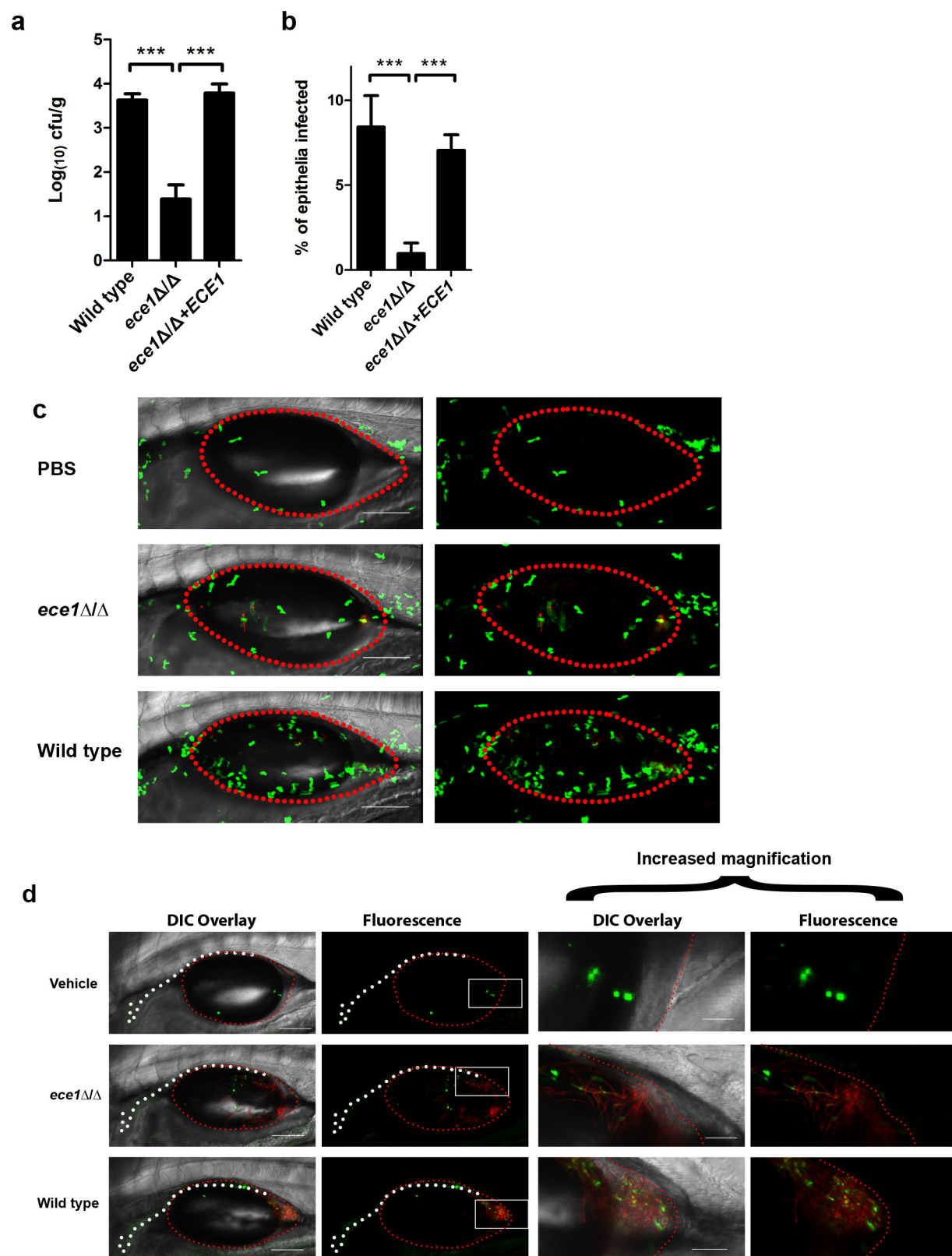
32. Rupniak, H. T. *et al.* Characteristics of four new human cell lines derived from squamous cell carcinomas of the head and neck. *J. Natl. Cancer Inst.* **75**, 621–635 (1985).
33. Mayer, F. L. *et al.* The novel *Candida albicans* transporter Dur31 is a multi-stage pathogenicity factor. *PLoS Pathog.* **8**, e1002592 (2012).
34. Gillum, A. M., Tsay, E. Y. & Kirsch, D. R. Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF* mutations. *Mol. Gen. Genet.* **198**, 179–182 (1984).
35. Citiulo, F. *et al.* *Candida albicans* scavenges host zinc via Pra1 during endothelial invasion. *PLoS Pathog.* **8**, e1002777 (2012).
36. Gola, S., Martin, R., Walther, A., Dunkler, A. & Wendland, J. New modules for PCR-based gene targeting in *Candida albicans*: rapid and efficient gene targeting using 100 bp of flanking homology region. *Yeast* **20**, 1339–1347 (2003).
37. Wilson, R. B., Davis, D. & Mitchell, A. P. Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. *J. Bacteriol.* **181**, 1868–1874 (1999).
38. Walther, A. & Wendland, J. An improved transformation protocol for the human fungal pathogen *Candida albicans*. *Curr. Genet.* **42**, 339–343 (2003).
39. Murad, A. M., Lee, P. R., Broadbent, I. D., Barelle, C. J. & Brown, A. J. Clp10, an efficient and convenient integrating vector for *Candida albicans*. *Yeast* **16**, 325–327 (2000).
40. Gratacap, R. L., Rawls, J. F. & Wheeler, R. T. Mucosal candidiasis elicits NF- κ B activation, proinflammatory gene expression and localized neutrophilia in zebrafish. *Dis. Model. Mech.* **6**, 1260–1270 (2013).
41. Moyes, D. L. *et al.* A biphasic innate immune MAPK response discriminates between the yeast and hyphal forms of *Candida albicans* in epithelial cells. *Cell Host Microbe* **8**, 225–235 (2010).
42. Moyes, D. L. *et al.* *Candida albicans* yeast and hyphae are discriminated by MAPK signaling in vaginal epithelial cells. *PLoS ONE* **6**, e26580 (2011).
43. Wächter, B., Wilson, D., Haedicke, K., Dalle, F. & Hube, B. From attachment to damage: defined genes of *Candida albicans* mediate adhesion, invasion and damage during interaction with oral epithelial cells. *PLoS ONE* **6**, e17046 (2011).
44. Trinh, L. A. *et al.* A versatile gene trap to visualize and interrogate the function of the vertebrate proteome. *Genes Dev.* **25**, 2306–2320 (2011).
45. Solis, N. V. & Filler, S. G. Mouse model of oropharyngeal candidiasis. *Nature Protocols* **7**, 637–642 (2012).
46. Cranfield, C., Carne, S., Martinac, B. & Cornell, B. The assembly and use of tethered bilayer lipid membranes (tBLMs). *Methods Mol. Biol.* **1232**, 45–53 (2015).
47. Cranfield, C. G. *et al.* Transient potential gradients and impedance measures of tethered bilayer lipid membranes: pore-forming peptide insertion and the effect of electroporation. *Biophys. J.* **106**, 182–189 (2014).
48. Schromm, A. B. *et al.* Lipopolysaccharide-binding protein mediates CD14-independent intercalation of lipopolysaccharide into phospholipid membranes. *FEBS Lett.* **399**, 267–271 (1996).
49. Chen, G. C. & Yang, J. T. 2-Point calibration of circular dichrometer with D-10-camphorsulfonic acid. *Anal. Lett.* **10**, 1195–1207 (1977).
50. Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc. Natl Acad. Sci. USA* **69**, 3561–3566 (1972).
51. Gutsmann, T., Heimbürg, T., Keyser, U., Mahendran, K. R. & Winterhalter, M. Protein reconstitution into freestanding planar lipid membranes for electrophysiological characterization. *Nature Protocols* **10**, 188–198 (2015).
52. Gillum, A. M., Tsay, E. Y. & Kirsch, D. R. Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF* mutations. *Mol. Gen. Genet.* **198**, 179–182 (1984).
53. Wilson, R. B., Davis, D. & Mitchell, A. P. Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. *J. Bacteriol.* **181**, 1868–1874 (1999).
54. Fonzi, W. A. & Irwin, M. Y. Isogenic strain construction and gene mapping in *Candida albicans*. *Genetics* **134**, 717–728 (1993).
55. Davis, D., Wilson, R. B. & Mitchell, A. P. *RLM101*-dependent and-independent pathways govern pH responses in *Candida albicans*. *Mol. Cell. Biol.* **20**, 971–978 (2000).
56. Braun, B. R. & Johnson, A. D. *TUP1*, *CPH1* and *EFG1* make independent contributions to filamentation in *Candida albicans*. *Genetics* **155**, 57–67 (2000).
57. Lo, H. J. *et al.* Nonfilamentous *C. albicans* mutants are avirulent. *Cell* **90**, 939–949 (1997).
58. Moyes, D. L. *et al.* A biphasic innate immune MAPK response discriminates between the yeast and hyphal forms of *Candida albicans* in epithelial cells. *Cell Host Microbe* **8**, 225–235 (2010).
59. Zakikhany, K. *et al.* In vivo transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination. *Cell. Microbiol.* **9**, 2938–2954 (2007).
60. Cao, F. *et al.* The Flo8 transcription factor is essential for hyphal development and virulence in *Candida albicans*. *Mol. Biol. Cell* **17**, 295–307 (2006).
61. Bockmühl, D. P., Krishnamurthy, S., Gerads, M., Sonneborn, A. & Ernst, J. F. Distinct and redundant roles of the two protein kinase A isoforms Tpk1p and Tpk2p in morphogenesis and growth of *Candida albicans*. *Mol. Microbiol.* **42**, 1243–1257 (2001).
62. Sonneborn, A. *et al.* Protein kinase A encoded by *TPK2* regulates dimorphism of *Candida albicans*. *Mol. Microbiol.* **35**, 386–396 (2000).
63. Palmer, G. E., Cashmore, A. & Sturtevant, J. *Candida albicans* *VPS11* is required for vacuole biogenesis and germ tube formation. *Eukaryot. Cell* **2**, 411–421 (2003).
64. Zou, H., Fang, H. M., Zhu, Y. & Wang, Y. *Candida albicans* Cyr1, Cap1 and G-actin form a sensor/effecter apparatus for activating cAMP synthesis in hyphal growth. *Mol. Microbiol.* **75**, 579–591 (2010).
65. Bates, S. *et al.* Outer chain N-glycans are required for cell wall integrity and virulence of *Candida albicans*. *J. Biol. Chem.* **281**, 90–98 (2006).
66. Murciano, C. *et al.* *Candida albicans* cell wall glycosylation may be indirectly required for activation of epithelial cell proinflammatory responses. *Infect. Immun.* **79**, 4902–4911 (2011).
67. Newport, G. & Agabian, N. *KEX2* influences *Candida albicans* proteinase secretion and hyphal formation. *J. Biol. Chem.* **272**, 28954–28961 (1997).
68. Murad, A. M. *et al.* *NRG1* represses yeast-hypha morphogenesis and hypha-specific gene expression in *Candida albicans*. *EMBO J.* **20**, 4742–4752 (2001).
69. Liu, H., Kohler, J. & Fink, G. R. Suppression of hyphal formation in *Candida albicans* by mutation of a *STE12* homolog. *Science* **266**, 1723–1726 (1994).
70. Lane, S., Zhou, S., Pan, T., Dai, Q. & Liu, H. The basic helix-loop-helix transcription factor Cph2 regulates hyphal development in *Candida albicans* partly via *TEC1*. *Mol. Cell. Biol.* **21**, 6418–6428 (2001).
71. White, S. J. *et al.* Self-regulation of *Candida albicans* population size during GI colonization. *PLoS Pathog.* **3**, e184 (2007).
72. Brown, D. H., Jr, Giusani, A. D., Chen, X. & Kumamoto, C. A. Filamentous growth of *Candida albicans* in response to physical environmental cues and its regulation by the unique *CZF1* gene. *Mol. Microbiol.* **34**, 651–662 (1999).
73. Kadosh, D. & Johnson, A. D. Rfg1, a protein related to the *Saccharomyces cerevisiae* hypoxic regulator Rox1, controls filamentous growth and virulence in *Candida albicans*. *Mol. Cell. Biol.* **21**, 2496–2505 (2001).
74. San José, C., Monge, R. A., Perez-Diaz, R., Pla, J. & Nombela, C. The mitogen-activated protein kinase homolog *HOG1* gene controls glycerol accumulation in the pathogenic fungus *Candida albicans*. *J. Bacteriol.* **178**, 5850–5852 (1996).
75. Firon, A. *et al.* The *SUN41* and *SUN42* genes are essential for cell separation in *Candida albicans*. *Mol. Microbiol.* **66**, 1256–1275 (2007).
76. de Boer, A. D. *et al.* The *Candida albicans* cell wall protein Rhd3/Pga29 is abundant in the yeast form and contributes to virulence. *Yeast* **27**, 611–624 (2010).
77. Mühlischlegel, F. A. & Fonzi, W. A. *PHR2* of *Candida albicans* encodes a functional homolog of the pH-regulated gene *PHR1* with an inverted pattern of pH-dependent expression. *Mol. Cell. Biol.* **17**, 5960–5967 (1997).
78. Martin, R. *et al.* A core filamentation response network in *Candida albicans* is restricted to eight genes. *PLoS ONE* **8**, e58613 (2013).
79. Birse, C. E., Irwin, M. Y., Fonzi, W. A. & Sypherd, P. S. Cloning and characterization of *ECE1*, a gene expressed in association with cell elongation of the dimorphic pathogen *Candida albicans*. *Infect. Immun.* **61**, 3648–3655 (1993).
80. Navarro-García, F., Sanchez, M., Pla, J. & Nombela, C. Functional characterization of the *MKC1* gene of *Candida albicans*, which encodes a mitogen-activated protein kinase homolog related to cell integrity. *Mol. Cell. Biol.* **15**, 2197–2206 (1995).

81. Hausauer, D. L., Gerami-Nejad, M., Kistler-Anderson, C. & Gale, C. A. Hyphal guidance and invasive growth in *Candida albicans* require the Ras-like GTPase Rsr1p and its GTPase-activating protein Bud2p. *Eukaryot. Cell* **4**, 1273–1286 (2005).
82. Sentandreu, M., Elorza, M. V., Sentandreu, R. & Fonzi, W. A. Cloning and characterization of *PRA1*, a gene encoding a novel pH-regulated antigen of *Candida albicans*. *J. Bacteriol.* **180**, 282–289 (1998).
83. Pardini, G. *et al.* The *CRH* family coding for cell wall glycosylphosphatidylinositol proteins with a predicted transglycosidase domain affects cell wall organization and virulence of *Candida albicans*. *J. Biol. Chem.* **281**, 40399–40411 (2006).
84. Braun, B. R., Head, W. S., Wang, M. X. & Johnson, A. D. Identification and characterization of *TUP1*-regulated genes in *Candida albicans*. *Genetics* **156**, 31–44 (2000).
85. Fradin, C. *et al.* Granulocytes govern the transcriptional response, morphology and proliferation of *Candida albicans* in human blood. *Mol. Microbiol.* **56**, 397–415 (2005).
86. Staab, J. F., Bradway, S. D., Fidel, P. L. & Sundstrom, P. Adhesive and mammalian transglutaminase substrate properties of *Candida albicans* Hwp1. *Science* **283**, 1535–1538 (1999).
87. Bailey, D. A., Feldmann, P. J., Bovey, M., Gow, N. A. & Brown, A. J. The *Candida albicans* *HYR1* gene, which is activated in response to hyphal development, belongs to a gene family encoding yeast cell wall proteins. *J. Bacteriol.* **178**, 5353–5360 (1996).
88. Sandini, S., La Valle, R., De Bernardis, F., Macri, C. & Cassone, A. The 65 kDa mannoprotein gene of *Candida albicans* encodes a putative β -glucanase adhesin required for hyphal morphogenesis and experimental pathogenicity. *Cell. Microbiol.* **9**, 1223–1238 (2007).
89. Csank, C. *et al.* Roles of the *Candida albicans* mitogen-activated protein kinase homolog, Cek1p, in hyphal development and systemic candidiasis. *Infect. Immun.* **66**, 2713–2721 (1998).
90. Hube, B. *et al.* Disruption of each of the secreted aspartyl proteinase genes *SAP1*, *SAP2*, and *SAP3* of *Candida albicans* attenuates virulence. *Infect. Immun.* **65**, 3529–3538 (1997).
91. Taylor, B. N. *et al.* Induction of *SAP7* correlates with virulence in an intravenous infection model of candidiasis but not in a vaginal infection model in mice. *Infect. Immun.* **73**, 7061–7063 (2005).
92. Schild, L. *et al.* Proteolytic cleavage of covalently linked cell wall proteins by *Candida albicans* Sap9 and Sap10. *Eukaryot. Cell* **10**, 98–109 (2011).
93. Zhao, X. *et al.* *ALS3* and *ALS8* represent a single locus that encodes a *Candida albicans* adhesin; functional comparisons between Als3p and Als1p. *Microbiol.* **150**, 2415–2428 (2004).
94. Murciano, C. *et al.* Evaluation of the role of *Candida albicans* agglutinin-like sequence (Als) proteins in human oral epithelial cell interactions. *PLoS ONE* **7**, e33362 (2012).
95. Zhao, X., Oh, S. H., Yeater, K. M. & Hoyer, L. L. Analysis of the *Candida albicans* Als2p and Als4p adhesins suggests the potential for compensatory function within the Als family. *Microbiol.* **151**, 1619–1630 (2005).
96. Zhao, X., Oh, S. H. & Hoyer, L. L. Deletion of *ALS5*, *ALS6* or *ALS7* increases adhesion of *Candida albicans* to human vascular endothelial and buccal epithelial cells. *Med. Mycol.* **45**, 429–434 (2007).
97. Zhao, X., Oh, S. H. & Hoyer, L. L. Unequal contribution of *ALS9* alleles to adhesion between *Candida albicans* and human vascular endothelial cells. *Microbiol.* **153**, 2342–2350 (2007).
98. Timpel, C., Strahl-Bolsinger, S., Ziegelbauer, K. & Ernst, J. F. Multiple functions of Pmt1p-mediated protein O-mannosylation in the fungal pathogen *Candida albicans*. *J. Biol. Chem.* **273**, 20837–20846 (1998).
99. Bates, S. *et al.* *Candida albicans* Pmr1p, a secretory pathway P-type $\text{Ca}^{2+}/\text{Mn}^{2+}$ -ATPase, is required for glycosylation and virulence. *J. Biol. Chem.* **280**, 23408–23415 (2005).
100. Hobson, R. P. *et al.* Loss of cell wall mannosylphosphate in *Candida albicans* does not influence macrophage recognition. *J. Biol. Chem.* **279**, 39628–39635 (2004).
101. Southard, S. B., Specht, C. A., Mishra, C., Chen-Weiner, J. & Robbins, P. W. Molecular analysis of the *Candida albicans* homolog of *Saccharomyces cerevisiae* *MNN9*, required for glycosylation of cell wall mannoproteins. *J. Bacteriol.* **181**, 7439–7448 (1999).
102. Munro, C. A. *et al.* Mnt1p and Mnt2p of *Candida albicans* are partially redundant α -1,2-mannosyltransferases that participate in O-linked mannosylation and are required for adhesion and virulence. *J. Biol. Chem.* **280**, 1051–1060 (2005).
103. Mio, T. *et al.* Role of three chitin synthase genes in the growth of *Candida albicans*. *J. Bacteriol.* **178**, 2416–2419 (1996).
104. Mille, C. *et al.* Inactivation of *CaMIT1* inhibits *Candida albicans* phospholipomannan β -mannosylation, reduces virulence, and alters cell wall protein β -mannosylation. *J. Biol. Chem.* **279**, 47952–47960 (2004).
105. Mille, C. *et al.* Identification of a new family of genes involved in β -1,2-mannosylation of glycans in *Pichia pastoris* and *Candida albicans*. *J. Biol. Chem.* **283**, 9724–9736 (2008).
106. Mille, C. *et al.* Members 5 and 6 of the *Candida albicans* *BMT* family encode enzymes acting specifically on β -mannosylation of the phospholipomannan cell-wall glycosphingolipid. *Glycobiol.* **22**, 1332–1342 (2012).
107. Mio, T. *et al.* Cloning of the *Candida albicans* homolog of *Saccharomyces cerevisiae* *GSC1/FKS1* and its involvement in β -1,3-glucan synthesis. *J. Bacteriol.* **179**, 4096–4105 (1997).
108. Mio, T. *et al.* Isolation of the *Candida albicans* homologs of *Saccharomyces cerevisiae* *KRE6* and *SKN1*: expression and physiological function. *J. Bacteriol.* **179**, 2363–2372 (1997).
109. Staab, J. F. & Sundstrom, P. *URA3* as a selectable marker for disruption and virulence assessment of *Candida albicans* genes. *Trends Microbiol.* **11**, 69–73 (2003).
110. Murad, A. M., Lee, P. R., Broadbent, I. D., Barelle, C. J. & Brown, A. J. Clp10, an efficient and convenient integrating vector for *Candida albicans*. *Yeast* **16**, 325–327 (2000).



Extended Data Figure 1 | *C. albicans* ECE1 expression and phenotypic effects of ECE1 gene deletion. **a**, Relative expression (vs $t = 0$) of ECE1 in *C. albicans* wild type over time after addition of yeast cells to TR146 epithelial cells as measured by RT-qPCR. **b**, Imaging confirmation of ECE1 expression over time within *C. albicans* wild-type strain. *C. albicans* cells expressing GFP under the control of the ECE1 5' intragenic region, containing the ECE1 promoter, were grown on TR146 epithelial cells and stained with calcofluor white (CFW, post-permeabilization) to show cell wall chitin and Alexa-Fluor-647-labelled concanavalin A (ConA, pre-permeabilization) to show carbohydrates. A composite image showing CFW, ConA, GFP and the brightfield (BF) image is shown. **c**, Scanning electron micrographs (top panels, 5 h) and light microscopy (bottom panels, 24 h) showing no gross abnormalities in hypha formation between *C. albicans* wild-type (BWP17 + Clp30), ECE1-deletion (*ece1* Δ/Δ) and ECE1 re-integrand (*ece1* Δ/Δ + ECE1) strains after infection of TR146

epithelial cells. **d**, No difference in adhesion of *C. albicans* wild-type, *ece1* Δ/Δ and *ece1* Δ/Δ + ECE1 strains to TR146 epithelial cells after 60 min. **e**, No difference in invasion of *C. albicans* wild-type, *ece1* Δ/Δ and *ece1* Δ/Δ + ECE1 strains into TR146 epithelial cells after 3 h. **f**, Fluorescence staining of *C. albicans* wild-type and *ece1* Δ/Δ hyphae invading through TR146 epithelial cells. Fungal cells are stained with calcofluor white (CFW, post-permeabilization) and Alexa-Fluor-647-labelled concanavalin A (ConA, pre-permeabilization) to show cell wall chitin and carbohydrates, respectively, and to distinguish between invading hyphae (only stained after permeabilization) and non-invading hyphae (stained both pre- and post-permeabilization). Levels of chitin and β -glucan are comparable in both strains. White arrows indicate invasion into epithelial cells. Data shown are representative (**b**, **c**, **f**) or the mean (**a**, **d**, **e**) of three biological replicates. Error bars show \pm s.e.m.



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | *C. albicans* Ece1p is critical for mucosal virulence *in vivo*. **a**, Fungal burdens recovered from the tongues of mice infected with *C. albicans* wild-type (BWP17 + Clp30) (number of mice (n) = 13), *ECE1*-deletion (*ece1*Δ/Δ) (n = 20) and *ECE1* re-integrand (*ece1*Δ/Δ + *ECE1*) (n = 24) strains after a 2-day oropharyngeal infection. **b**, Average percentage of the entire tongue epithelium area infected in different groups of mice infected with the different *C. albicans* strains. **c**, Confocal imaging of 4-day post-fertilization (dpf) *mpo-gfp* transgenic zebrafish swimbladders infected with *C. albicans* wild-type (BWP17 + Clp30 + *dTomato*), *ECE1*-deletion (*ece1*Δ/Δ + *dTomato*) and *ECE1* re-integrand (*ece1*Δ/Δ + *ECE1* + *dTomato*) strains for 24 h. *C. albicans* cells appear red while neutrophils appear green. Red dots outline the swimbladder. Images are composites of maximum projections in the red and green channels (25 slices each, approximately 100 μm depth) with (left) or without (right) a single slice in the DIC channel overlay. Scale bars

represent 100 μm. **d**, Confocal imaging of 4 dpf zebrafish swimbladders infected with *C. albicans* wild-type (BWP17 + Clp30 + *dTomato*), *ECE1*-deletion (*ece1*Δ/Δ + *dTomato*) and *ECE1* re-integrand (*ece1*Δ/Δ + *ECE1* + *dTomato*) strains for 24 h stained with the fluorescent exclusion dye Sytox Green. *C. albicans* cells appear red and damaged epithelial cells appear green. White dots outline the pronephros and red dots outline the swimbladder. Images are composites of maximum projections in the red and green channels (25 slices each, approximately 100 μm depth) with (left) or without (right) a single slice in the DIC channel overlay. High magnification images of the white boxes are shown. Scale bars (bottom right) represent 100 μm (low magnification) and 30 μm (high magnification). Data shown are the mean (**a**, **b**) or representative (**c**, **d**) of at least three biological replicates. Error bars show ± s.e.m. Data were analysed by Mann–Whitney *U*-test. *** P < 0.001.

a

***Candida albicans* Ece1p amino acid sequence:**

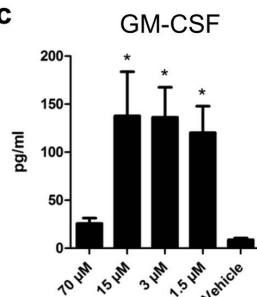
MKFSKIACATVFALSSQAAIIHHAPEFNMKR DVAPAAPAPADQAPTVPAPQEFNTAITKR SIIGIIMGILGNIPQVIQIIMSIVKA
FKGNKR EDIDSVVAGI IADMPFVVRVAVDTAMTSVASTKR DGANDDVANAVVRLPEIVARVATGVQQSIENAKR DGVDPVGLNLVANA
PRLISNVFDGSETVQQA KR DGLDFLDELLQRLPQLITRSAESALKDSQPV KR DAGSVALSNLIKSIETVGIENAAQIVSERDIS
SLIEEYFGKA



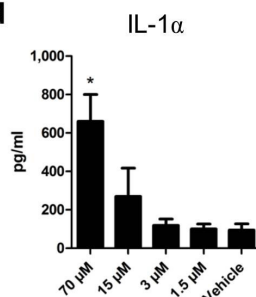
b

Ece1-I₁₋₃₁ MKFSKIACATVFALSSQAAIIHHAPEFNMKR
Ece1-II₃₂₋₆₁ DVAPAAPAPADQAPTVPAPQEFNTAITKR
Ece1-III₆₂₋₉₃ SIIGIIMGILGNIPQVIQIIMSIVKAFKGNKR
Ece1-IV₉₄₋₁₂₆ EDIDSVVAGI IADMPFVVRVAVDTAMTSVASTKR
Ece1-V₁₂₇₋₁₆₀ DGANDDVANAVVRLPEIVARVATGVQQSIENAKR
Ece1-VI₁₆₁₋₁₉₄ DGVDPVGLNLVANA PRLISNVFDGSETVQQA KR
Ece1-VII₁₉₅₋₂₂₈ DGLDFLDELLQRLPQLITRSAESALKDSQPV KR
Ece1-VIII₂₂₉₋₂₇₁ DAGSVALSNLIKSIETVGIENAAQIVSERDIS SLIEEYFGKA

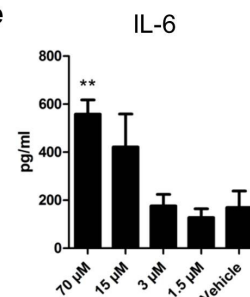
c



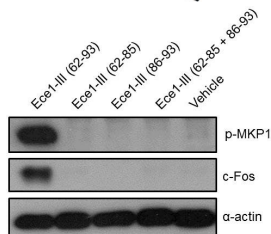
d



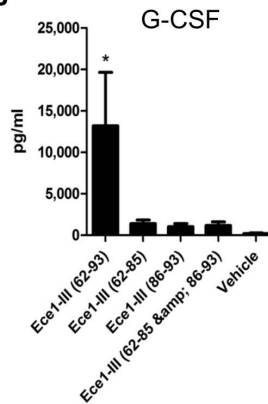
e



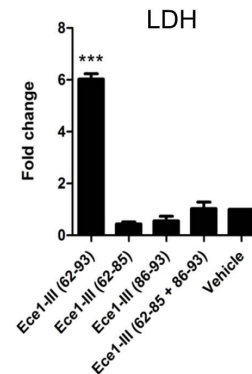
f



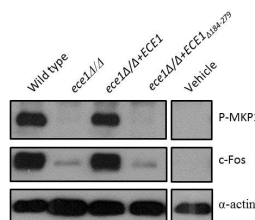
g



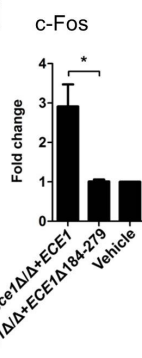
h



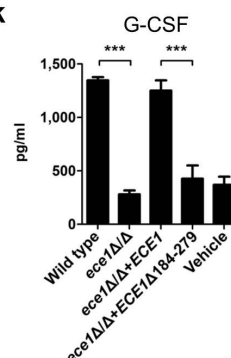
i



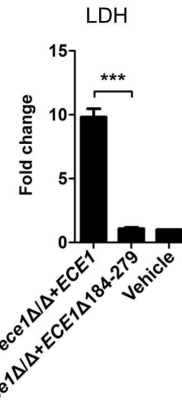
j



k



l

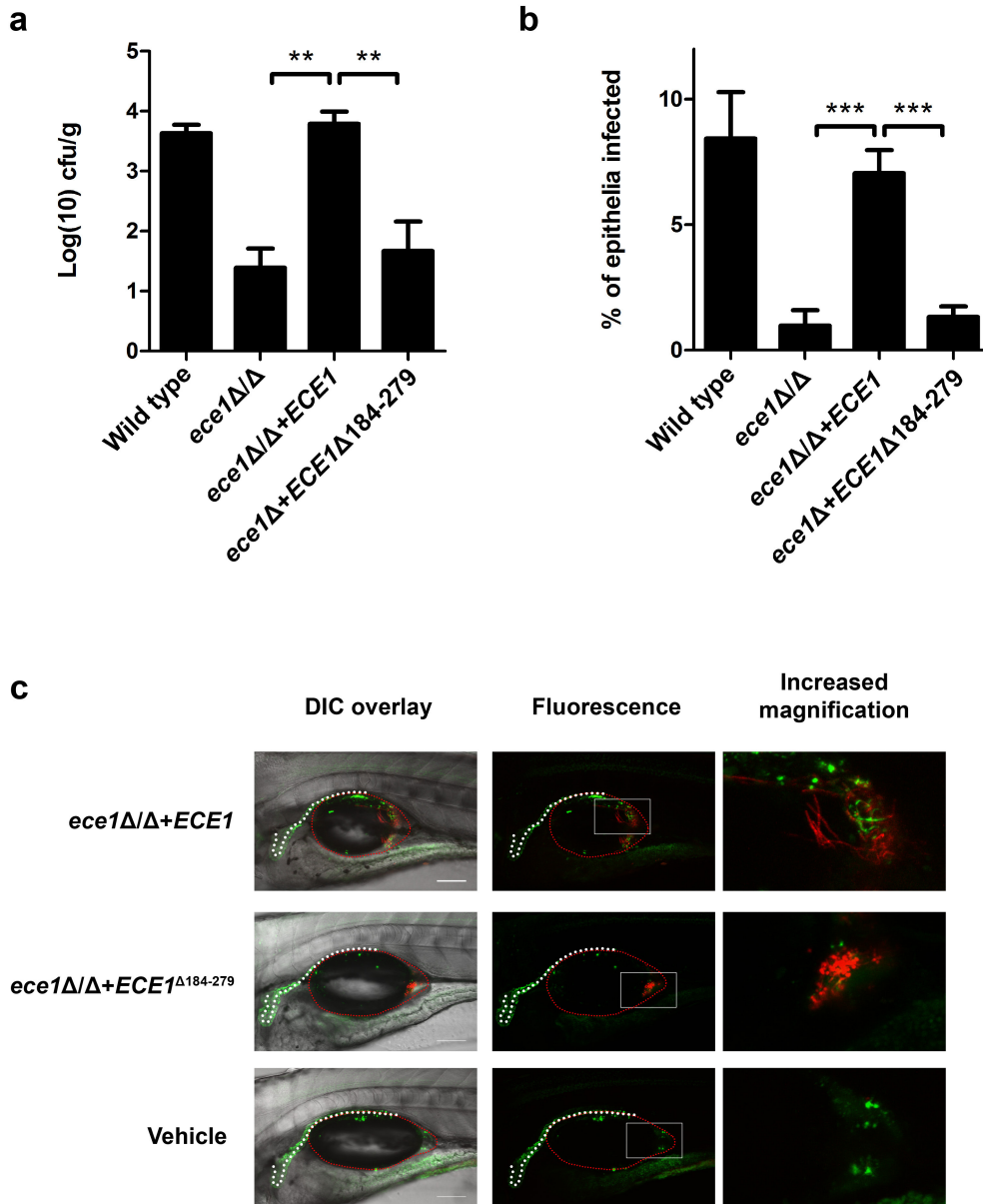


Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Ece1-III₆₂₋₉₃ is the active region of Ece1p.

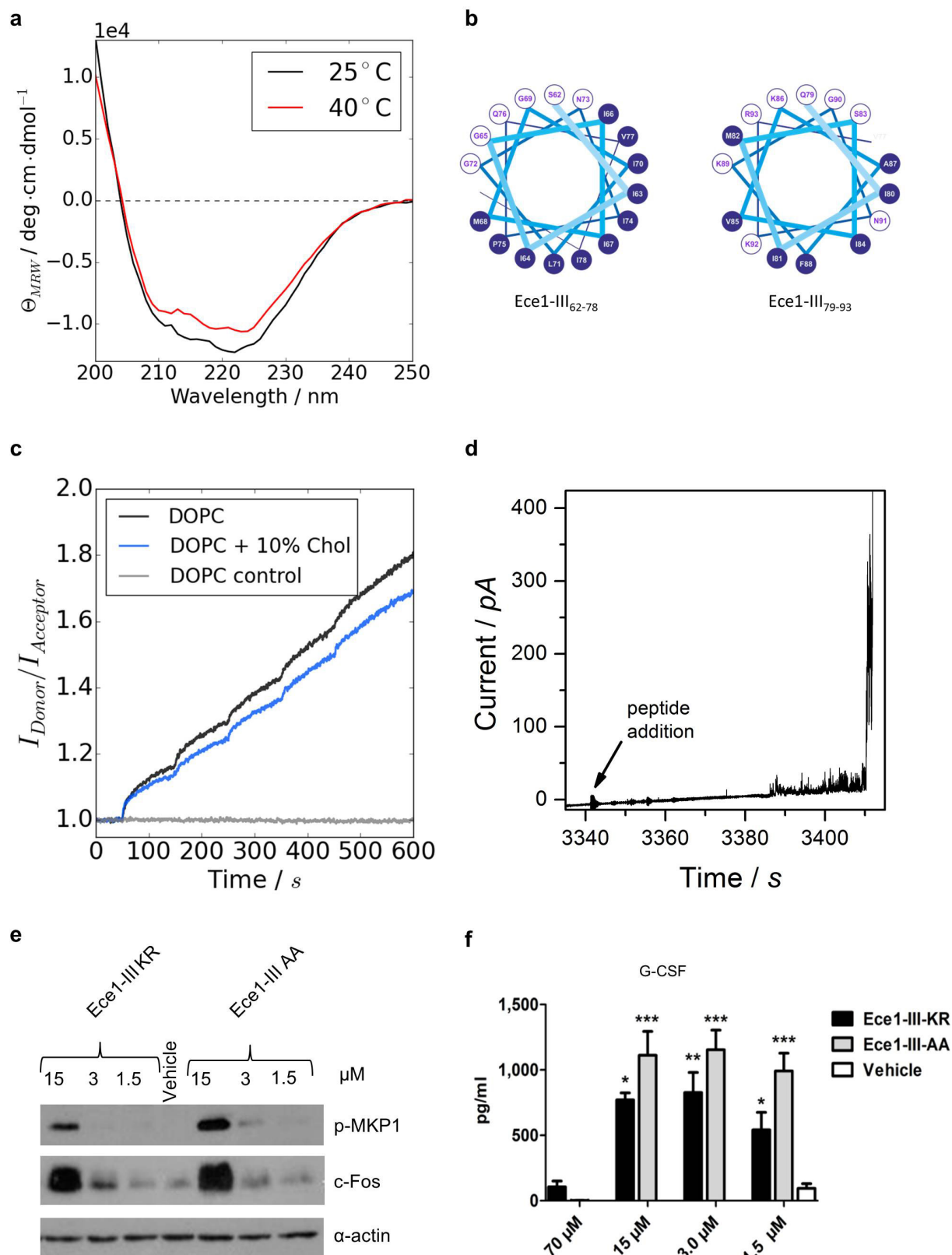
a, Amino acid sequence of Ece1p and a schematic of the protein, indicating the signal peptide (SP), lysine-arginine motifs (KR) at the C terminus of each peptide, and the processed peptides (Ece1-I–VIII) produced by Kex2p cleavage. **b**, Amino acid sequences of the processed peptides (Ece1-I–VIII) produced by Kex2p cleavage. **c–e**, Induction of GM-CSF (**c**), IL-1 α (**d**) and IL-6 (**e**) secreted after stimulation of TR146 epithelial cells for 24 h with varying concentrations of Ece1-III₆₂₋₉₃ (1.5–70 μ M). **f**, Phosphorylation of MKP-1 and c-Fos production after 2 h treatment of TR146 epithelial cells with 15 μ M of Ece1-III₆₂₋₈₅ (hydrophobic region), Ece1-III₈₆₋₉₃ (hydrophilic region), Ece1-III₆₂₋₈₅ and Ece1-III₈₆₋₉₃ together, or Ece1-III₆₂₋₉₃ alone. **g**, Induction of G-CSF secretion after 24 h treatment of TR146 epithelial cells with 15 μ M of Ece1-III₆₂₋₈₅, Ece1-III₈₆₋₉₃,

Ece1-III₆₂₋₈₅ and Ece1-III₈₆₋₉₃ together, or Ece1-III₆₂₋₉₃ alone. **h**, Fold change induction of LDH release after 24 h treatment of TR146 epithelial cells with 70 μ M of Ece1-III₆₂₋₈₅, Ece1-III₈₆₋₉₃, Ece1-III₆₂₋₈₅ and Ece1-III₈₆₋₉₃ together, or Ece1-III₆₂₋₉₃ alone. **i**, Induction of p-MKP-1 and c-Fos 2 h post-infection (p.i.) with the indicated *C. albicans* strains (MOI = 10). **j**, c-Fos DNA binding induction 3 h p.i. with indicated *C. albicans* strains (MOI = 10). **k**, G-CSF secretion and **l**, LDH release 24 h p.i. with indicated *C. albicans* strains (MOI = 0.01). Data shown are representative (**f**, **i**) or the mean (**c–e**, **g**, **h**, **j–l**) of three biological replicates. Error bars show \pm s.e.m. Data were analysed by one-way ANOVA (**c–e**, **g**, **h**, **k–l**) or *t*-test (**j**). * P < 0.05, ** P < 0.01, *** P < 0.001 (compared with vehicle control). For gel image, see Supplementary Fig. 1.



Extended Data Figure 4 | Ece1-III₆₂₋₉₃ is required for *C. albicans* mucosal infection. **a**, Fungal burdens recovered from the tongues of mice infected with *C. albicans* wild-type (BWP17 + Clp30) (number of mice (n) = 13), *ECE1*-deletion (*ece1* Δ/Δ) (n = 20), *ECE1* re-integrand (*ece1* Δ/Δ + *ECE1*) (n = 24) and Ece1-III₆₂₋₉₃ deletion (*ece1* Δ/Δ + *ECE1* $\Delta_{184-279}$) (n = 10) strains after 2-day oropharyngeal infection. **b**, Average percentage of the entire tongue epithelium area infected in different groups of mice infected with the different *C. albicans* strains. **c**, Confocal imaging of 4 dpf zebrafish swimbladders infected with *C. albicans* Ece1-III₆₂₋₉₃ deletion (*ece1* Δ/Δ + *ECE1* $\Delta_{184-279}$ + *dTomato*)

and *ECE1* re-integrand (*ece1* Δ/Δ + *ECE1* + *dTomato*) strains for 24 h stained with the fluorescent exclusion dye Sytox Green. *C. albicans* cells appear red and damaged cells appear green. White dots outline the pronephros and red dots outline the swimbladder. Images are composites of maximum projections in the red and green channels (25 slices each, approximately 100 μ m depth) with (left) or without (right) a single slice in the DIC channel overlay. Scale bars (bottom right) represent 100 μ m. Data shown are the mean (**a**, **b**) or representative (**c**), of at least three biological replicates. Error bars show \pm s.e.m. Data were analysed by Mann-Whitney *U*-test. ** P < 0.01, *** P < 0.001.

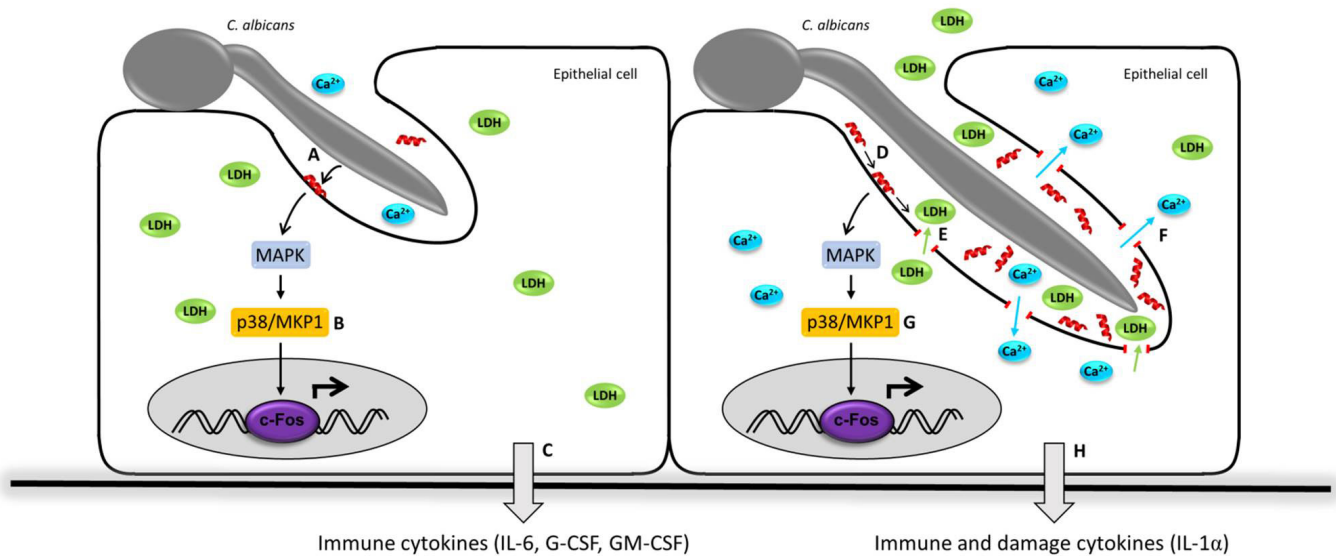


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Ece1-III₆₂₋₉₃ is a cytolytic α -helical peptide.

a, Circular dichroism spectra showing the α -helical conformation of Ece1-III₆₂₋₉₃ in buffer (100 mM KCl, 5 mM HEPES, pH 7). Increasing the temperature from 25 °C to 40 °C did not affect the stability of the α -helical structure. **b**, Diagram to illustrate the amphipathic nature of Ece1-III₆₂₋₉₃ (residues 62–78, left panel; residues 79–93, right panel). Residues with hydrophobic or polar/charged side chains are displayed with a blue and white background, respectively. Modified from output generated in PEPWHEEL (<http://emboss.bioinformatics.nl/cgi-bin/emboss/pepwheel>). **c**, Förster resonance energy transfer (FRET) experiments show the intercalation of Ece1-III₆₂₋₉₃ into lipid liposomes (10 μ M) composed of DOPC in the absence or presence of cholesterol. Peptide titration of

Ece1-III₆₂₋₉₃ to liposomes showed slightly enhanced intercalation for pure DOPC. **d**, Ece1-III₆₂₋₉₃ induced the permeabilization of planar lipid membranes composed of DOPC. The graph shows heterogeneous and transient lesions leading finally to a rupture of the membrane. Ece1-III₆₂₋₉₃ concentration was 0.125 μ M. **e**, Induction of p-MKP-1 and c-Fos 2 h in TR146 cells post stimulation (p.s.) with Ece1-III_{62-93KR} or Ece1-III_{62-93AA}. **f**, Secretion of G-CSF from TR146 cells 24 h p.s. with Ece1-III_{62-93KR} or Ece1-III_{62-93AA}. Data shown are representative (**a–e**) or mean (**f**) of at least three biological replicates. Error bars show \pm s.e.m. Data were analysed by one-way ANOVA (**f**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (compared with vehicle control). For gel images, see Supplementary Fig. 1.

Early stage infection/Sub-lytic Ece1-III concentration**Late stage infection/Lytic Ece1-III concentration**

Extended Data Figure 6 | Schematic of the role of Ece1-III in *C. albicans* infection of epithelial cells. a–h, During early stage infection of the mucosal surface by *C. albicans*, Ece1-III (red α -helix) is secreted into the invasion pocket created by the invading hypha (a), Sub-lytic concentrations of Ece1-III trigger epithelial signal transduction through MAPK, p38/MKP-1 and c-Fos (b), resulting in the production of immune regulatory cytokines (c), As the severity of the infection increases, Ece1-III accumulates (d), and once lytic concentrations are reached, causes

membrane damage and the release of lactate dehydrogenase from the host epithelium (e), concomitant with calcium influx (f). Epithelial signal transduction is maintained (g), and additionally induces the release of damage associated cytokines, such as IL-1 α (h). Ece1-III may also have activity on the epithelial surface outside of the invasion pocket and on neighbouring cells not in contact with hyphae if Ece1-III is produced in sufficient concentrations.

Extended Data Table 1 | *C. albicans* strains used in this study

Strain name	Strain/Gene Function	Strain Reference	Morphology [*]	Phospho-MKP1 [†]	c-Fos [†]	Cytokines [‡]	Damage [§]	Phenotype Reference
Controls								
SC5314	Wild type	[52]	Hyphae	Yes	Yes	Yes	Yes	This study
BWP17 & Clp30	Parental strain	[53]	Hyphae	Yes	Yes	Yes	Yes	This study
CAI-4 & Clp10	Parental strain	[54]	Hyphae	Yes	Yes	Yes	Yes	This study
CAF2-1	Parental strain	[54]	Hyphae	Yes	Yes	Yes	Yes	This study
DAY286	Parental strain	[55]	Hyphae	Yes	Yes	Yes	Yes	This study
Yeast-locked								
<i>efg1Δ/Δ</i>	Transcription factor	[56]	Yeast	No	No	No	No	This study
<i>efg1/cph1Δ/Δ</i>	Transcription factor/ Transcription factor	[57]	Yeast	No	No	No	No	[58]/This study
<i>eed1Δ/Δ</i>	RNA polymerase II regulator	[59]	Yeast	No	No	No	No	[58]/This study
<i>flo8Δ/Δ</i>	Transcription factor	[60]	Yeast	No	No	No	No	This study
<i>tpk1Δ/Δ</i>	cAMP-dependent protein kinase	[61]	Yeast	No	No	No	No	This study
<i>tpk2Δ/Δ</i>	cAMP-dependent protein kinase	[62]	Yeast	No	No	No	No	This study
<i>vps11Δ/Δ</i>	Protein trafficking	[63]	Yeast	No	No	No	No	This study
<i>cap1Δ/Δ</i>	Transcription factor	[64]	Yeast	No	No	Yes	No	This study
<i>och1Δ/Δ</i>	Alpha-1,6-mannosyltransferase	[65]	Yeast	No	No	No	No	[66]
<i>kex2Δ/Δ</i>	Processing enzyme	[67]	Yeast	No	No	No	No	This study
Hypha-producing								
<i>nrg1Δ/Δ</i>	Transcriptional corepressor	[68]	Hyphae	Yes	Yes	Yes	Yes	[58]/This study
<i>cph1Δ/Δ</i>	Transcription factor	[69]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>cph2Δ/Δ</i>	Transcription factor	[70]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>efh1Δ/Δ</i>	Transcription factor	[71]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>czf1Δ/Δ</i>	Transcription factor	[72]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>rfg1Δ/Δ</i>	Transcriptional repressor	[73]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>hog1Δ/Δ</i>	MAP kinase	[74]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>sun42Δ/Δ</i>	Adhesin-like protein	[75]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>pga29Δ/Δ</i>	GPI-anchored yeast-associated protein	[76]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>phr2Δ/Δ</i>	Glycosidase	[77]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>pga36Δ/Δ</i>	GPI-anchored protein	[78]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>ece1Δ/Δ</i>	Hypha-associated protein	[79]	Hyphae	No	No	No	No	This study
<i>mkc1Δ/Δ</i>	MAP kinase	[80]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>bud2Δ/Δ</i>	GTPase activating protein	[81]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>pra1Δ/Δ</i>	Zinc binding protein	[82]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>utr2/crh11/crh12Δ/Δ</i>	Putative wall glycosidase/transglycosylase	[83]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>wap1Δ/Δ</i>	Surface antigen on hyphae/buds	[84]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>sod5Δ/Δ</i>	Superoxide dismutase	[85]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>hwp1Δ/Δ</i>	Adhesin	[86]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>rbt1Δ/Δ</i>	Putative GPI-modified cell wall protein	[84]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>rbt5Δ/Δ</i>	Heme binding	[84]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>hyr1Δ/Δ</i>	GPI-anchored hyphal cell wall protein	[87]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>mp65Δ/Δ</i>	Cell surface mannoprotein	[88]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>cek1Δ/Δ</i>	ERK-family protein kinase	[89]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>sap2Δ/Δ</i>	Secreted aspartyl protease	[90]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>sap7Δ/Δ</i>	Secreted aspartyl protease	[91]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>sap9/sap10Δ/Δ</i>	Secreted aspartyl proteases	[92]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>als1Δ/Δ</i>	Agglutinin-like sequence protein	[93]	Hyphae	Yes	Yes	Yes	Yes	[94]
<i>als2Δ/Δ</i> <i>P_{MAL}-ALS2</i>	Agglutinin-like sequence protein	[95]	Hyphae	Yes	Yes	Yes	Yes	[94]
<i>als3Δ/Δ</i>	Adhesin	[93]	Hyphae	Yes	Yes	Partial [¶]	Partial [¶]	[94]
<i>als4Δ/Δ</i>	Agglutinin-like sequence protein	[95]	Hyphae	Yes	Yes	Yes	Yes	[94]
<i>als5Δ/Δ</i>	Agglutinin-like sequence protein	[96]	Hyphae	Yes	Yes	Yes	Yes	[94]
<i>als6Δ/Δ</i>	Agglutinin-like sequence protein	[96]	Hyphae	Yes	Yes	Yes	Yes	[94]
<i>als7Δ/Δ</i>	Agglutinin-like sequence protein	[96]	Hyphae	Yes	Yes	Yes	Yes	[94]
<i>als9Δ/Δ</i>	Agglutinin-like sequence protein	[97]	Hyphae	Yes	Yes	Yes	Yes	[94]
<i>pmt1Δ/Δ</i>	Mannosyltransferase	[98]	Hyphae	Partial [¶]	Partial [¶]	Partial [¶]	Partial [¶]	[66]
<i>pmr1Δ/Δ</i>	Secretory pathway ATPase	[99]	Hyphae	Partial [¶]	Partial [¶]	Partial [¶]	Partial [¶]	[66]
<i>mnn4Δ/Δ</i>	Regulator of mannosylphosphorylation	[100]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>mnn9Δ/Δ</i>	Putative mannosyltransferase	[101]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>mmt1/mmt2Δ/Δ</i>	Mannosyltransferases	[102]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>chs2/chs3Δ/Δ</i>	Chitin synthase/ Chitin synthase	[103]	Hyphae	Yes	Yes	Yes	Yes	This study
<i>mit1Δ/Δ</i>	Mannose:Inositolphosphoceramide mannosyltransferase	[104]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>bmt1Δ/Δ</i>	Beta-mannosyltransferase	[105]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>bmt2Δ/Δ</i>	Putative beta-mannosyltransferase	[105]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>bmt3Δ/Δ</i>	Beta-mannosyltransferase	[105]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>bmt4Δ/Δ</i>	Beta-mannosyltransferase	[105]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>bmt5Δ/Δ</i>	Putative beta-mannosyltransferase	[106]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>bmt6Δ/Δ</i>	Beta-mannosyltransferase	[106]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>gsc1Δ/GSC1</i>	Beta-1,3-glucan synthase catalytic subunit	[107]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>gs11Δ/Δ</i>	Beta-1,3-glucan synthase subunit	[107]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>gs12Δ/Δ</i>	Beta-1,3-glucan synthase subunit	[107]	Hyphae	Yes	Yes	Yes	Yes	[66]
<i>kre6Δ/KRE6</i>	Beta-1,6-glucan synthase subunit	[108]	Hyphae	Yes	Yes	Yes	Yes	This study

References 52–108 are cited in this table.

^{*}Morphology recorded 2 h post-infection on TR146 buccal cell monolayers; hyphae includes pseudohyphae.[†]Data based on western blotting.[‡]Cytokines include IL-1 α , IL-6 and G-CSF.[§]Damage measured by LDH assay.^{||}New *ece1Δ/Δ* also created in this study (see Extended Data Table 2). Original mutant (in red) produced by ref. 27 using the URA-blaster protocol³. A set of *ece1* mutants, including partial deletion of *ECE1* and a revertant, was produced in this study in the same genetic background using strain BWP17 to avoid a URA3 effect based on genomic location^{109,110}.[¶]Partial activation is due to lack of adhesion.

Extended Data Table 2 | *C. albicans* mutant strains constructed and used in this study

Strain description	Strain name	Genotype
BWP17+Clp30	M1477	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>RPS1/rps1::(URA3-HIS1-ARG4)</i>
<i>ece1Δ/Δ</i>	M2057	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>ece1::HIS1/ece1::ARG4</i> <i>RPS1/rps1::URA3</i>
<i>ece1Δ/Δ+ECE1</i>	M2059	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>ece1::HIS1/ece1::ARG4</i> <i>RPS1/rps1::(URA3-ECE1)</i>
<i>ece1Δ/Δ+ECE1_{Δ184-279}</i>	M2174	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>ece1::HIS1/ece1::ARG4</i> <i>RPS1/rps1::(URA3-ECE1^{Δ184-279})</i>
<i>kex1Δ/Δ</i>	M2258	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>kex1::HIS1/kex1::ARG4</i> <i>RPS1/rps1::URA3</i>
SC5314+ <i>pECE1-GFP</i> (<i>ECE1</i> promoter-GFP)	CA58	<i>ECE1/ece1::GFP-SAT1</i>
BWP17+Clp30+ <i>pENO1-dTom</i> (<i>ENO1</i> promoter-dTom)	RWC83	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>RPS1/rps1::(URA3-HIS1-ARG4)</i> <i>ENO1/eno1::dTom-SAT1</i>
<i>ece1Δ/Δ+pENO1-dTom</i> (<i>ENO1</i> promoter-dTom)	RWC84	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>ece1::HIS1/ece1::ARG4</i> <i>RPS1/rps1::URA3</i> <i>ENO1/eno1::dTom-SAT1</i>
<i>ece1Δ/Δ+ECE1</i> + dTomato	RWC85	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>ece1::HIS1/ece1::ARG4</i> <i>RPS1/rps1::(URA3-ECE1)</i> <i>ENO1/eno1::dTomato-NAT^r</i>
<i>ece1Δ/Δ+ECE1_{Δ184-279}</i> + dTomato	RWC86	<i>ura3::λimm434/ura3::λimm434</i> <i>iro1::λimm434/iro1::λimm434</i> <i>his1::hisG/his1::hisG</i> <i>arg4::hisG/arg4::hisG</i> <i>ece1::HIS1/ece1::ARG4</i> <i>RPS1/rps1::(URA3-ECE1^{Δ184-279})</i> <i>ENO1/eno1::dTomato-NAT^r</i>

Extended Data Table 3 | LC-MS/MS analysis of *C. albicans* Ece1-III

Ece1-III sequence	PSM Value* (% total Ece1-III [†]) (% total Ece1p [‡])				
	Wild Type	<i>ece1Δ/Δ+ECE1</i>	TR146+Wild type	rEce1p+rKex2p	<i>kex1Δ/Δ</i>
SIIGIIMGILGNIPQVIQIIMSIVKAFKGNK	699 (86%) (41%)	477 (89%) (35%)	79 (97.5%) (97.5%)	n/d [§]	49 (13.3%) (3.6%)
SIIGIIMGILGNIPQVIQIIMSIVKAFKGNKR	1 (0.1%) (0.06%)	1 (0.2%) (0.07%)	2 (2.5%) (2.5%)	248 (80%) (1.5%)	291 (78.9%) (21%)

*The number of peptide spectrum matches. Data for *ece1Δ/Δ* and *ece1Δ/Δ+ECE1_{Δ184-279}* are not included as no Ece1-III peptides were detected in either strain, as expected.

[†]Percentage of SIIGIIMGILGNIPQVIQIIMSIVKAFKGNK or SIIGIIMGILGNIPQVIQIIMSIVKAFKGNKR detected amongst all Ece1-III peptides found by LC-MS/MS.

[‡]Percentage of SIIGIIMGILGNIPQVIQIIMSIVKAFKGNK or SIIGIIMGILGNIPQVIQIIMSIVKAFKGNKR detected amongst all Ece1p peptides found by LC-MS/MS.

[§]n/d; not detected.

Extended Data Table 4 | Oligonucleotide primers used in this study

Primer name	Application	Sequence (5'-3')	Description
ECE1-FG	PCR	atcaataaccacattttcaaaattgtttttttttttatctctacaaca aacaactttctttttttactaccaactattttcattcgtaaagaagcttc gtacgctgcaggtc	Construction of <i>ECE1</i> deletion construct
ECE1-RG	PCR	cacaaaaacaacaattaaaaaatcagttacagcaaaagtgtcacaag acttatggaataaaagattaaagcttggtgaaaacaaattttatctgctgag cattctgatcatcgatgaattcgag	Construction of <i>ECE1</i> deletion construct
ECE1-RecF3k	PCR	gcacgcgtctaaagtgaggtaacaac	Construction of <i>ECE1</i> complementation plasmid
ECE1-RecR	PCR	ggtcgacccagacgttggtgc	Construction of <i>ECE1</i> complementation plasmid
ECE1-F1	PCR	ggcttctcataaatgaaggctcg	Confirmation of <i>ECE1</i> deletion
ECE1-R1	PCR	gccgaatcaatctgtctgccac	Confirmation of <i>ECE1</i> deletion
KEX1-FG	PCR	tatcttttttttttttttttaccatcttcatattttacaacccctgatacctt acctaacaacacacatctatctttaatcaaacacaaatcaattgaa gcttcgtacgctgcaggtc	Construction of <i>KEX1</i> deletion construct
KEX1-RG	PCR	tcacaatctagattattgtagggtgtatagacaaaaataaaatcaaaact attattgttataaaatctacaagatcttcaatctcactgtaccgaaaaat tctgatcatcgatgaattcgag	Construction of <i>KEX1</i> deletion construct
KEX1-F1	PCR	ggaaagccataagaattgga	Confirmation of <i>KEX1</i> deletion
KEX1-R1	PCR	aggaaagctgtggtgtagtg	Confirmation of <i>KEX1</i> deletion
HIS-F2	PCR	ggacgaattgaagaagctggtgcaaccg	Confirmation of <i>ECE1/KEX1</i> deletion
HIS-R2	PCR	caacgaaatggcctcccctaccacag	Confirmation of <i>ECE1/KEX1</i> deletion
ARG-F2	PCR	ggatatgttgctactgatttag	Confirmation of <i>ECE1/KEX1</i> deletion
ARG-R2	PCR	aatggatcagtgaggcaccggtg	Confirmation of <i>ECE1/KEX1</i> deletion
ECE1-Flnt1	PCR	ctaacgtttttgatggcgtcctgg	Confirmation of plasmid integration
URAF2	PCR	ggagttggattagatgataaaggatgatgg	Confirmation of plasmid integration
RPF-1	PCR	gagcaggtgtacacacacacatcttg	Confirmation of plasmid integration
RPF-2	PCR	cgccaaagagtttccctattatc	Confirmation of plasmid integration
Pep3-F1	PCR	gaagatatcgattctgttggctgg	Excision of Ece1-III ₆₂₋₉₃ from <i>ECE1</i>
Pep3-R1	PCR	cagaatcgatatcttcttcttggtaatagcagattgaattcttg	Excision of Ece1-III ₆₂₋₉₃ from <i>ECE1</i>
5'ECE1prom-NarI	PCR	gatcggcgcctccagccactattttgtacctgt	Amplification of <i>ECE1</i> promoter region for <i>ECE1</i> promoter-GFP construct
3'ECE1prom-XhoI	PCR	tcagctcaggtttaacgaatggaaaatagttggtag	Amplification of <i>ECE1</i> promoter region for <i>ECE1</i> promoter-GFP construct
5'ECE1term-SacII	PCR	gatcccgccgagcagataaaaaattgtttccacaag	Amplification of <i>ECE1</i> terminator region for <i>ECE1</i> promoter-GFP construct
5'ECE1term-SacI	PCR	tcaggagctccgttaagaatatgaatgacagattggtc	Amplification of <i>ECE1</i> terminator region for <i>ECE1</i> promoter-GFP construct
G1-ECE1	PCR	ctcgtgattagagttcaagagt	Confirmation of <i>ECE1</i> -GFP plasmid integration (5' end)
GFP veri rev	PCR	tgatctgggtatctcgaaagcat	Confirmation of <i>ECE1</i> -GFP plasmid integration (5' end)
G4-ECE1	PCR	tggaaagattcacttgagttggaac	Confirmation of <i>ECE1</i> -GFP plasmid integration (3' end)
X3-SAT1	PCR	gtgaagtgtgaaggggag	Confirmation of <i>ECE1</i> -GFP plasmid integration (3' end)
pENO1 FW	PCR	tccttggctggcactgaactcg	Confirmation of pENO1-dTom plasmid integration
dTom REV	PCR	aaggctctaccttcacctcacc	Confirmation of pENO1-dTom plasmid integration
ACT1-F	qPCR	tcagaccagctgatttaggttg	Quantification of actin cDNA
ACT1-R	qPCR	gtgaacaatggatggaccag	Quantification of actin cDNA
ECE1-F	qPCR	atcgaaaatccaagagag	Quantification of <i>ECE1</i> cDNA
ECE1-R	qPCR	agcattttcaatccgacag	Quantification of <i>ECE1</i> cDNA

Recent near-Earth supernovae probed by global deposition of interstellar radioactive ^{60}Fe

A. Wallner¹, J. Feige^{2†}, N. Kinoshita³, M. Paul⁴, L. K. Fifield¹, R. Golser², M. Honda⁵, U. Linnemann⁶, H. Matsuzaki⁷, S. Merchel⁸, G. Rugel⁸, S. G. Tims¹, P. Steier², T. Yamagata⁹ & S. R. Winkler²

The rate of supernovae in our local Galactic neighbourhood within a distance of about 100 parsecs from Earth is estimated to be one every 2–4 million years, based on the total rate in the Milky Way (2.0 ± 0.7 per century^{1,2}). Recent massive-star and supernova activity in Earth's vicinity may be traced by radionuclides with half-lives of up to 100 million years^{3–6}, if trapped in interstellar dust grains that penetrate the Solar System. One such radionuclide is ^{60}Fe (with a half-life of 2.6 million years)^{7,8}, which is ejected in supernova explosions and winds from massive stars^{1,2,9}. Here we report that the ^{60}Fe signal observed previously in deep-sea crusts^{10,11} is global, extended in time and of interstellar origin from multiple events. We analysed deep-sea archives from all major oceans for ^{60}Fe deposition via the accretion of interstellar dust particles. Our results reveal ^{60}Fe interstellar influxes onto Earth at 1.5–3.2 million years ago and at 6.5–8.7 million years ago. The signal measured implies that a few per cent of fresh ^{60}Fe was captured in dust and deposited on Earth. Our findings indicate multiple supernova and massive-star events during the last ten million years at distances of up to 100 parsecs.

The density and temperature distribution of the interstellar medium (ISM) is highly variable, with typical substructures of about 50–150 pc (superbubbles) having lifetimes of some ten million years (Myr). Several supernova explosions over the last 14 Myr or so shaped the present structure of the local superbubble (the Local Bubble)^{12–14}. The Solar System, now embedded in the Local Bubble, is expected to have faced fronts of supernova ejecta and accumulated material from massive

stars. To enter the Solar System, any material from the ISM must be condensed into larger dust grains to avoid being deflected away by the solar wind and interplanetary magnetic field^{3,10,11}. ISM dust particles were indeed identified at Earth orbit¹⁵ and may accumulate on Earth in archives such as deep-sea sediments and ferromanganese (FeMn) crusts and nodules, which retain time information over millions of years. ^{60}Fe as well as ^{26}Al (half-life $t_{1/2} = 0.71$ Myr) have been observed^{1,9} in the ISM as a result of many supernovae and emission from massive stars. Direct detection of 'live' radionuclides^{3–5,10,11} on Earth would provide insight into recent and nearby nucleosynthesis in massive stars^{14,16,17}, dust formation and transport into the Solar System. Extraterrestrial ^{60}Fe has in fact already been observed in FeMn crusts in pioneering studies at TU Munich^{10,11}, and is interpreted as being of supernova^{10,11,18} or (micro)meteoritic origin^{19,20}.

Here, we determined the ^{60}Fe contents of three different deep-sea archives (four sediment cores, two FeMn crusts and two FeMn nodules) recovered from the Indian, Pacific and Atlantic oceans respectively (Supplementary Fig. 1). All were dated via their ^{10}Be (half-life $t_{1/2} = 1.39$ Myr) content, complemented by ^{26}Al for the sediments²¹. All radionuclides (^{60}Fe , ^{26}Al and ^{10}Be) were counted using accelerator mass spectrometry (AMS) (Supplementary Information). The sediment cores provided a record from 1.7–3.2 Myr ago with a time resolution of <30,000 years (30 kyr), bracketed by recent and ~5–7-Myr-old samples. The Pacific 'Crust-1' extends from the present to 10.9 Myr ago with ~2.2-Myr time resolution and 'Crust-2' extends

Table 1 | Averaged $^{60}\text{Fe}/\text{Fe}$ atom ratios from AMS measurements at ANU of the sediment samples

Sediment cores	Sediment samples	Time period (Myr)	^{60}Fe counts detected	$^{60}\text{Fe}/\text{Fe}$ (10^{-15} atoms per atom)	Background- and decay-corrected $^{60}\text{Fe}/\text{Fe}$ (10^{-15} atoms per atom)	Fe concentration (10^{-2} grams per gram)	^{60}Fe concentration (10^4 atoms per gram)	^{60}Fe deposition rates (atoms $\text{cm}^{-2} \text{yr}^{-1}$)	^{60}Fe deposition (10^6 atoms cm^{-2} per layer)
45–21/50–02	5	<0.2	2	0.06 ± 0.04	0.02 ± 0.02	0.30 ± 0.10	<0.2	<0.2	NA
49–53/45–21	14	1.71–2.0	123	1.67 ± 0.15	2.52 ± 0.23	0.23 ± 0.01	6.0 ± 0.6	22.8 ± 2.3	6.5 ± 0.7
49–53/45–21/50–02	11	2.0–2.3	51	1.51 ± 0.21	2.48 ± 0.35	0.24 ± 0.01	6.7 ± 1.0	24.8 ± 3.6	7.4 ± 1.1
49–53/45–21/50–02	7	2.3–2.6	33	1.96 ± 0.34	3.50 ± 0.61	0.17 ± 0.01	6.5 ± 1.2	27.1 ± 5.0	8.1 ± 1.5
49–53/45–16	7	2.6–2.9	54	3.40 ± 0.46	6.61 ± 0.90	0.16 ± 0.01	10.3 ± 1.5	34.8 ± 5.2	10.4 ± 1.5
49–53/45–16	6	2.9–3.18	27	1.18 ± 0.23	2.41 ± 0.47	0.13 ± 0.01	3.4 ± 0.7	11.4 ± 2.4	3.0 ± 0.6
45–16	2	~4–7*	1	0.11 ± 0.11	0.20 ± 0.30	0.14 ± 0.01	<0.4	<1	NA
Commercial iron	99	Background	7	0.042 ± 0.015	NA	NA	NA	NA	NA

52 sediment samples from four sediment cores (Eltanin) from the Indian Ocean were analysed (individual data are listed in the Supplementary Information) as well as a series of blank samples (commercial iron). For the measured $^{60}\text{Fe}/\text{Fe}$ ratios, the uncertainties from ^{60}Fe denote statistical uncertainties only (1σ , using Poisson statistics). For background- and decay-corrected $^{60}\text{Fe}/\text{Fe}$ data, surface and old layers are compatible with the measurement background obtained from chemistry blank samples. The age is based on the ^{26}Al and ^{10}Be data and tie-points of magnetic reversals. The stable iron content of the leached material was measured via inductively coupled plasma mass spectrometry (ICP-MS) (averaged values). The average leachable Fe content for the four cores was measured to 0.18% (core Eltanin 49–53), 0.25% (45–16), 0.3% (45–21) and 0.45% (50–02). The mean dry density of the sediments was 1.16 g cm^{-3} . For sediments with 100% incorporation efficiency, the Fe deposition equals the terrestrial fluence.

*Uncertain by ~1 Myr.

NA, not applicable.

¹Department of Nuclear Physics, Research School of Physics and Engineering, The Australian National University (ANU), Canberra, Australian Capital Territory 2601, Australia. ²University of Vienna, Faculty of Physics—Isotope Research, VERA Laboratory, Währinger Straße 17, 1090 Vienna, Austria. ³Institute of Technology, Shimizu Corporation, Tokyo 135-8530, Japan. ⁴Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. ⁵Graduate School of Pure and Applied Sciences, University of Tsukuba, Ibaraki 305-8577, Japan. ⁶Senckenberg Collections of Natural History Dresden, GeoPlasmaLab, Königsbrücker Landstraße 159, Dresden 01109, Germany. ⁷MALT (Micro Analysis Laboratory, Tandem accelerator), The University Museum, The University of Tokyo, Tokyo 113-0032, Japan. ⁸Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute for Resource Technology, 01328 Dresden, Germany. ⁹Graduate School of Integrated Basic Sciences, Nihon University, Tokyo 156-8550, Japan. [†]Present address: Department of Astronomy and Astrophysics, Berlin Institute of Technology, Hardenbergstrasse 36, 10623 Berlin, Germany.

Table 2 | $^{60}\text{Fe}/\text{Fe}$ ratios from AMS measurements at ANU of the Pacific and Atlantic FeMn samples

Crust-1									
Layer	Depth (mm)	Time period (Myr)	^{60}Fe counts detected	$^{60}\text{Fe}/\text{Fe}$ (10^{-15} atoms per atom)	Background- and decay-corrected $^{60}\text{Fe}/\text{Fe}$ (10^{-15} atoms per atom)	Fe concentration (10^{-2} grams per gram)	^{60}Fe concentration (10^6 atoms per gram)	^{60}Fe incorporation rates (atoms $\text{cm}^{-2} \text{yr}^{-1}$)	^{60}Fe incorporation (10^6 atoms cm^{-2} per layer)
Layer 1	0–5	0–2.17	23	0.96 ± 0.25	1.19 ± 0.33	12.8 ± 0.1	1.64 ± 0.45	0.72 ± 0.20	1.56 ± 0.43
Layer 2	5–10	2.17–4.35	74	1.58 ± 0.23	3.60 ± 0.55	11.8 ± 0.1	4.56 ± 0.68	2.00 ± 0.30	4.34 ± 0.65
Layer 3	10–15	4.35–6.52	2	0.14 ± 0.12	0.36 ± 0.51	11.4 ± 0.1	0.34 ± 0.62	0.19 ± 0.27	0.40 ± 0.59
Layer 4	15–20	6.52–8.70	26	0.38 ± 0.09	2.49 ± 0.97	13.7 ± 0.2	3.67 ± 1.43	1.61 ± 0.63	3.49 ± 1.36
Layer 5	20–25	8.70–10.87	3	0.15 ± 0.11	1.20 ± 1.50	11.2 ± 0.2	1.45 ± 1.82	0.64 ± 0.80	1.38 ± 1.73
Total for 0–4.35 Myr and 6.52–8.70 Myr			123	NA	NA	NA	NA	NA	9.4 ± 0.9
Crust-2									
Layers 1, 2	0–1.0	1.20–1.41	5	1.06 ± 0.48	1.50 ± 0.67	11.6 ± 0.1	1.87 ± 0.83	1.67 ± 0.75	0.36 ± 0.16
Layers 3–5	1.0–2.5	1.41–1.73	37	1.28 ± 0.21	2.01 ± 0.33	11.9 ± 0.1	2.60 ± 0.43	2.33 ± 0.38	0.74 ± 0.12
Layers 6–8	2.5–4.0	1.73–2.05	32	0.84 ± 0.15	1.39 ± 0.25	11.8 ± 0.1	1.73 ± 0.31	1.55 ± 0.27	0.50 ± 0.09
Layers 9, 10	4.0–5.8	2.05–3.05	20	0.55 ± 0.12	1.09 ± 0.24	11.1 ± 0.1	1.35 ± 0.30	0.46 ± 0.10	0.46 ± 0.10
Layers 11, 12	5.8–7.7	3.05–4.11	2	0.12 ± 0.08	0.29 ± 0.21	10.4 ± 0.1	0.37 ± 0.26	0.13 ± 0.09	0.13 ± 0.09
Layers 13–26	7.7–21.0	4.11–7	1	<0.03	<0.1	9	<0.1	<0.09	<0.06
Total for 1.20–3.05 Myr			94	NA	NA	NA	NA	NA	2.19 ± 0.22
Nodule 21									
Layer 1	0–3	0–1.8	3	0.16 ± 0.11	<0.23	15 ± 2	<0.4	<0.13	<0.23
Layer 2	3–6	1.8–3.3	13	0.47 ± 0.16	0.60 ± 0.22	15 ± 2	0.97 ± 0.36	0.40 ± 0.15	0.55 ± 0.21
Layers 3, 4	6–17	3.3–5.4	5	0.18 ± 0.06	<0.10	15 ± 2	<0.16	<0.1	<0.23
Total for 1.8–3.3 Myr			13	0.47 ± 0.16	0.60 ± 0.22	15 ± 2	0.97 ± 0.36	0.40 ± 0.15	0.55 ± 0.21
Nodule 24									
Layer 1	0–4	0–1.9	15	0.56 ± 0.18	0.71 ± 0.23	15 ± 2	1.13 ± 0.37	0.34 ± 0.11	0.86 ± 0.41
Layer 2	4–8	1.9–3.3	5	0.23 ± 0.12	0.45 ± 0.23	15 ± 2	0.71 ± 0.37	0.27 ± 0.15	0.54 ± 0.34
Layers 3, 4	8–19	3.3–5.4	1	0.05 ± 0.05	<0.15	15 ± 2	<0.24	<0.1	<0.25
Total for 0–3.3 Myr			20	0.40 ± 0.10	0.58 ± 0.13	15 ± 2	0.92 ± 0.21	0.31 ± 0.09	1.40 ± 0.50

The table shows data for layered samples of the two Pacific FeMn crust samples (Crust-1 and Crust-2) and of the two Atlantic FeMn nodules (numbers 21 and 24). For measured $^{60}\text{Fe}/\text{Fe}$ ratios, the uncertainties from ^{60}Fe denote statistical uncertainties only (1σ , using Poisson statistics). For background- and decay-corrected $^{60}\text{Fe}/\text{Fe}$ data, the machine background of $^{60}\text{Fe}/\text{Fe} = (0.042 \pm 0.015) \times 10^{-15}$ was used for the subsequent background correction. Ages are based on the ^{10}Be data and have an uncertainty of ± 0.3 Myr (Crust-1) and ± 0.5 Myr (Crust-2 and nodules). Stable iron content of the dissolved crust material and of the leachate in case of the nodules was measured via ICP-MS. The mean dry density of the crust and nodule samples was 1.9 g cm^{-3} . The Fe concentration listed for layers 13–26 of Crust-2 is the average; individual data range between 5.5% and 13.1%. 'Total' gives the sum of ^{60}Fe counts detected and the total ^{60}Fe incorporation of all layers that are significantly above the measurement background.

from 1.2–7 Myr ago (~ 100 -kyr resolution). Two nodules covered a time period from the present to 5.4 Myr ago (~ 2 -Myr resolution).

In the sediment, 288 ^{60}Fe -events were registered for the time period 1.71–3.18 Myr ago (45 individual samples) with a mean isotopic ratio of $^{60}\text{Fe}/\text{Fe} = (1.79 \pm 0.10) \times 10^{-15}$, a factor of ~ 40 above the measurement background of $(0.042 \pm 0.015) \times 10^{-15}$ (Table 1). None of the recent or old sediment samples show evidence for ^{60}Fe above background (three ^{60}Fe -events). The first two layers in Crust-1 gave ^{60}Fe -signals 4σ and 7σ above background; layers 3 and 5 are close to the measurement background, but layer 4, which spans the period 6.5–8.7 Myr ago, has a higher ratio ($\sim 4\sigma$ above background, Table 2). For Crust-2 a clear ^{60}Fe -signal was also found at <3.5 Myr. The nodules support this finding (Table 3, Supplementary Tables 3–5).

In summary, two clear ^{60}Fe signals with a total of 538 ^{60}Fe events were observed. In the sediments, the signal covers the time period 1.7–3.2 Myr. In the two crusts, ^{60}Fe is found up to 3.5 Myr ago and up to about 4 Myr ago, respectively, with a second influx between 6.5 Myr and 8.7 Myr ago. The nodules confirm the presence of ^{60}Fe at <3.3 Myr ago. No ^{60}Fe signal is found in recent (<0.2 Myr old) or older (≥ 5 Myr old) sediments and nodules, or in crust layers 4.4–6.5 Myr and 8.7–10.9 Myr old.

Between 1.7 Myr ago and 3.18 Myr ago, the ^{60}Fe deposition rate into the sediments was ~ 11 – 35 ^{60}Fe atoms $\text{cm}^{-2} \text{yr}^{-1}$ (300-kyr averages), whereas incorporation rates into crust material were much lower at 1–2 atoms $\text{cm}^{-2} \text{yr}^{-1}$ (Fig. 1; all data are decay-corrected). This suggests an incorporation efficiency into Crust-1 and Crust-2 of 17% and 7%, respectively. The deposition in the 1.5-Myr interval covered by the signal

in the sediment is $(35 \pm 2) \times 10^6$ atoms cm^{-2} . For the second ^{60}Fe signal (6.5–8.7 Myr ago, Crust-1, 17% incorporation efficiency) it is $(21 \pm 6) \times 10^6$ atoms cm^{-2} (Tables 2 and 3).

Although the 1.5-Myr time spread of ^{60}Fe influx measured in the present work exceeds the ~ 0.8 Myr previously reported for crust 237KD^{11,18}, the two time profiles are not inconsistent given the lower counting statistics and signal-to-background ratio in ref. 11. Furthermore, the marginally positive result for the same time period for an Atlantic sediment¹⁸ is consistent with our data, considering their higher sedimentation rates and stable Fe contents. ^{60}Fe has also been reported in lunar material, though without time information²², and recently in Pacific sediments^{23,31}.

Clearly, our data are incompatible with a constant ^{60}Fe production or deposition. A terrestrial origin can be ruled out, because there is no suitable target for cosmic-ray-induced production and anthropogenic input would be concentrated in the surface layer. Since ^{60}Fe was found in each of the major oceans, it is reasonable to assume a uniform global distribution. A micro-meteoritic or meteoritic origin can be excluded, given that the measured cosmic-dust flux is 400 times lower than would be required (Supplementary Information and Supplementary Fig. 6). Similarly, a hypothetical break-up of a single object, comparable to the asteroid invoked in relation to the Cretaceous–Tertiary event 65 Myr ago, would have delivered 4,500 times less ^{60}Fe .

We assume that the extraterrestrial ^{60}Fe flux through Earth's cross-section is homogeneously distributed over Earth's surface. Thus, the measured mean deposition of ~ 24.5 atoms $\text{cm}^{-2} \text{yr}^{-1}$ (1.7–3.2 Myr

Table 3 | Summary of ^{60}Fe deposition at various locations

Deep-sea archive	Cores	Location	Time period (Myr)	^{60}Fe detector events	^{60}Fe deposition (10^6 atoms cm^{-2})
Sediment	4	Indian Ocean	1.71–3.18	288	35.4 ± 2.6
FeMn Crust-1	2	Pacific Ocean	0–4.35	97	5.9 ± 0.8
FeMn Crust-1			6.52–8.70	26	3.5 ± 1.4
FeMn Crust-2			1.2–3.1	94	2.2 ± 0.2
FeMn nodules	2	Atlantic Ocean	1.8–3.3	13	0.6 ± 0.2
			0–3.3	20	1.4 ± 0.5
FeMn Mona Pihua ¹⁰	1	Pacific Ocean	0–5.9	21	$\sim 9^{+11}_{-6}$
FeMn 237KD ¹¹	1	Pacific Ocean	1.74–2.61*	69	$1.5 \pm 0.4^*$
Lunar material ²²	4	Moon	Integral	†	~ 10

Data were obtained in this work and as given in the literature^{10,11,22} (no correction for incorporation efficiency). Uncertainties are 1σ . For Crust-1 and Crust-2 an incorporation efficiency of 17% and 7%, respectively, has to be taken into account to calculate the ^{60}Fe fluence from the deposition values; similarly 2% and 4% for the nodules.

*Adjusted for revised ^{60}Fe and ^{10}Be half-life values^{7,8}.

†Not listed in ref. 22.

‡Background-corrected, adjusted for revised ^{60}Fe half-life and interpolation between the two layers.

signal) corresponds to a ^{60}Fe flux of $98 \text{ atoms cm}^{-2} \text{ yr}^{-1}$ into the inner Solar System or integrated over 1.5 Myr to an ^{60}Fe fluence of $(1.46 \pm 0.15) \times 10^8 \text{ atoms cm}^{-2}$ at Earth orbit; the fluence for the older event is $(1.2 \pm 0.4) \times 10^8 \text{ atoms cm}^{-2}$. Interstellar grains, filtered by the Solar System in size to an average of $\sim 0.5 \mu\text{m}$, were detected by space missions¹⁵, suggesting that $(6 \pm 3)\%$ of the mass of ISM dust reaches the inner Solar System⁶. These grains follow the flow velocity of the ISM.

Assuming that the ^{60}Fe -loaded grains follow the same mass distribution as determined for ISM grains at Earth orbit, we deduce an interstellar ^{60}Fe concentration in dust of $(2.8 \pm 1.4) \times 10^{-11} \text{ }^{60}\text{Fe atoms cm}^{-3}$ for 1.7–3.2 Myr and, integrated over the full period of 11 Myr, an average concentration of $\sim (5\text{--}15) \times 10^{-12} \text{ atoms cm}^{-3}$. Observations of ^{60}Fe decay^{1,9} and nucleosynthesis models² suggest an average Galaxy concentration of $\sim 6 \times 10^{-12} \text{ }^{60}\text{Fe atoms cm}^{-3}$ (Supplementary Information), in agreement with the 11-Myr period of local data reported here.

^{60}Fe is produced in massive stars^{2,24–27} in their late phases, predominantly just before supernova explosions, and then ejected into space. (Super)asymptotic-giant-branch stars also produce and eject ^{60}Fe through their stellar winds for a period of about 50 kyr, leading to a time profile similar to supernovae; however, their contribution to the Galactic ^{60}Fe inventory is small²⁸.

Models suggest a travel time of about 200 kyr with a time spread of approximately 100–400 kyr (ref. 5) for ejecta from a single supernova at a distance from Earth of about 100 pc. Our measured spread of about 1.5 Myr is inconsistent with the interpretation in terms of ejecta from a single supernova (or asymptotic-giant-branch star) moving across the Solar System (Supplementary Fig. 6). It suggests multiple supernova and massive-star events within the last ten million years or so in Earth's vicinity and during two distinct periods 1.7–3.2 Myr ago and around 6.5–8.7 Myr ago. The recent time profile would be compatible with movement across the Solar System of ejecta in a series of supernova fronts in short succession within 1.5 Myr.

This would, however, require a high frequency of supernovae (around two to three supernovae per million years), since large fluctuations were not observed in the time profile. Alternatively, the ejecta containing the ^{60}Fe -bearing grains could have come to rest in the ambient ISM and diffused into volumes or clouds that were then traversed by the Solar System¹⁸.

The Solar System is currently embedded in a flow of ISM material with interstellar grains moving parallel to the flow of neutral interstellar

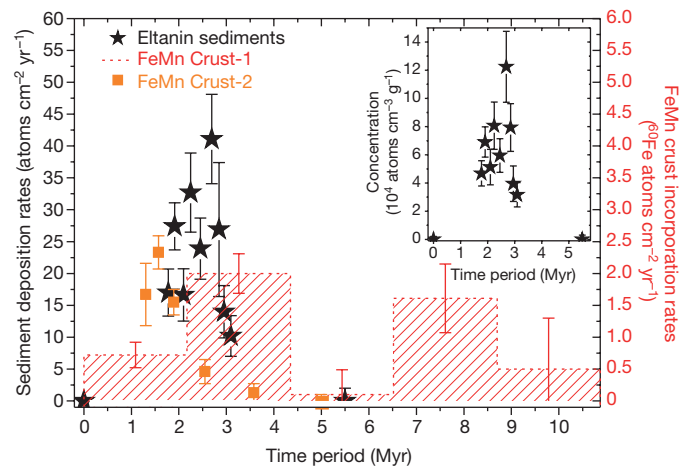


Figure 1 | Deposition rates for sediment (150-kyr averaged data) and incorporation rates for two crust samples. ^{60}Fe concentrations (^{60}Fe per gram) for the sediment are given in the inset; they were on average 6.7×10^4 atoms per gram between 1.7 Myr and 3.2 Myr, but 260×10^4 atoms per gram of crust and 95×10^4 atoms per gram of nodule, reflecting the difference in growth rate and incorporation efficiency (see Supplementary Information). The error bars (1σ Poisson statistics) include all uncertainties and scale with decay correction, so that uncertainties and upper limits become larger for older samples. The absolute ages for the sediment samples have an uncertainty of 0.1 Myr, except for the 5.5-Myr-old sediments, which have an uncertainty of about 1 Myr. The age of Crust-1 has an uncertainty of 0.3 Myr and the age of Crust-2 has an uncertainty of 0.5 Myr.

gas in local ISM clouds, which suggests a common history or driver²⁹. Such clouds have been suggested as part of an expanding superbubble shell driven by supernovae and winds from massive stars^{12–14,29}.

Assuming the ejecta originate from a distance of 70–100 pc (the approximate limit of the Local Bubble) and that ^{60}Fe is equally distributed into the outer shell of size 30 pc (a distance representing 1.5 Myr of travel), that is, assuming a spherical shell of mean radius 70–100 pc with a thickness of 30 pc, we deduce a total ^{60}Fe mass trapped in ISM dust of $(5\text{--}11) \times 10^{-5}$ solar masses (M_\odot) in the shell volume. This number represents a lower limit because it reflects the fraction of ^{60}Fe condensed into dust without correction for radioactive decay and neglects the granularity of clumpy ejecta. Models predict core-collapse and electron-capture supernova nucleosynthesis yields for ^{60}Fe to be $(0.5\text{--}14) \times 10^{-5} M_\odot$ for stars of mass $8 M_\odot\text{--}25 M_\odot$, depending on the progenitor mass, and with large uncertainties in the nuclear physics input^{2,24–27}. (Super)asymptotic-giant-branch stars produce $(0.003\text{--}1) \times 10^{-5} M_\odot \text{ }^{60}\text{Fe}$ ²⁸. Our observed signals therefore favour supernova events. The fraction of ^{60}Fe in dust can be roughly estimated by a comparison of our measured ^{60}Fe deposition with nucleosynthesis yields. Under these assumptions and assuming reasonable distances (20–100 pc) about 0.4% to 9% of ^{60}Fe would have been trapped in ISM dust particles (Supplementary Information, Supplementary Figs 7 and 8).

Comparing our data with a similar work for ISM ^{244}Pu in sediments and crust samples⁶ yields a $^{244}\text{Pu}/^{60}\text{Fe}$ atomic ratio of $< 2\text{--}3 \times 10^{-5}$ (2σ) during periods of elevated ^{60}Fe deposition over the past ten million years, which agrees with the recently reported low ^{244}Pu supernova yields⁶ (Supplementary Information).

This broad and global ^{60}Fe influx on Earth demonstrates recent (within the past ten million years) and widespread massive-star ejections in our near Galactic neighbourhood (less than 100 pc from Earth), most probably from supernova explosions. Interestingly, the older event coincides with a strong increase in ^3He and temperature change at about 8 Myr ago³⁰, while the more recent activity starting about 3 Myr ago occurred at the same time as Earth's temperature started to decrease during the Plio–Pleistocene transition.

Received 23 August 2015; accepted 28 January 2016.

1. Diehl, R. Nuclear astrophysics lessons from INTEGRAL. *Rep. Prog. Phys.* **76**, 026301 (2013).
2. Timmes, F. X. & Woosley, S. E. Gamma-ray line signals from ^{26}Al and ^{60}Fe in the galaxies of the Local Group. *Astrophys. J.* **481**, L81 (1997).
3. Ellis, J., Fields, B. D. & Schramm, D. N. Geological isotope anomalies as signatures of nearby supernovae. *Astrophys. J.* **470**, 1227–1236 (1996).
4. Korschinek, G., Faestermann, T., Knie, K. & Schmidt, C. ^{60}Fe , a promising AMS isotope for many applications. *Radiocarbon* **38**, 68 (1996).
5. Fry, B. J., Fields, B. D. & Ellis, J. R. Astrophysical shrapnel: discriminating among extra-Earth stellar explosion sources of live radioactive isotopes. *Astrophys. J.* **800**, 71 (2015).
6. Wallner, A. *et al.* Abundance of live ^{244}Pu in deep-sea reservoirs on Earth points to rarity of actinide nucleosynthesis. *Nature Commun.* **6**, 5956 (2015).
7. Rugel, G. *et al.* New measurement of the ^{60}Fe half-life. *Phys. Rev. Lett.* **103**, 072502 (2009).
8. Wallner, A. *et al.* Settling the half-life of ^{60}Fe —fundamental for a versatile astrophysical chronometer. *Phys. Rev. Lett.* **114**, 041101 (2015).
9. Wang, W. *et al.* SPI observations of the diffuse ^{60}Fe emission in the Galaxy. *Astron. Astrophys.* **469**, 1005–1012 (2007).
10. Knie, K. *et al.* Indication for supernova produced ^{60}Fe activity on Earth. *Phys. Rev. Lett.* **83**, 18–21 (1999).
11. Knie, K. *et al.* ^{60}Fe anomaly in a deep-sea manganese crust and implications for a nearby supernova source. *Phys. Rev. Lett.* **93**, 171103 (2004).
12. Maíz-Apellániz, J. The origin of the Local Bubble. *Astrophys. J.* **560**, L83–L86 (2001).
13. Breitschwerdt, D. & de Avillez, M. A. The history and future of the Local and Loop I bubbles. *Astron. Astrophys.* **452**, L1–L5 (2006).
14. Benítez, N., Maíz-Apellániz, J. & Canelles, M. Evidence for nearby supernova explosions. *Phys. Rev. Lett.* **88**, 081101 (2002).
15. Mann, I. Interstellar dust in the Solar System. *Annu. Rev. Astron. Astrophys.* **48**, 173–203 (2010).
16. Beech, M. The past, present and future supernova threat to Earth's biosphere. *Astrophys. Space Sci.* **336**, 287–302 (2011).
17. Ruderman, M. A. Possible consequences of nearby supernova explosions for atmospheric ozone and terrestrial life. *Science* **184**, 1079–1081 (1974).
18. Fitoussi, C. *et al.* Search for supernova-produced ^{60}Fe in a marine sediment. *Phys. Rev. Lett.* **101**, 121101 (2008).
19. Stuart, F. M. & Lee, M. R. Micrometeorites and extraterrestrial He in a ferromanganese crust from the Pacific Ocean. *Chem. Geol.* **322–323**, 209–214 (2012).
20. Basu, S., Stuart, F. M., Schnabel, C. & Klemm, V. Galactic-cosmic-ray-produced ^3He in a ferromanganese crust: any supernova ^{60}Fe excess on Earth? *Phys. Rev. Lett.* **98**, 141103 (2007).
21. Feige, J. *et al.* AMS measurements of cosmogenic and supernova-ejected radionuclides in deep-sea sediment cores. In *Eur. Phys. J. Web Conf.* **63**, 3003, <http://dx.doi.org/10.1051/epjconf/20136303003> (2013).
22. Fimiani, L. *et al.* Evidence for deposition of interstellar material on the lunar surface. *Lunar Planet. Sci. Conf.* **45**, 1778 (2014).
23. Ludwig, P. Search for ^{60}Fe of supernova origin in Earth's microfossil record. PhD thesis, TU Munich (2015).
24. Woosley, S. E. & Weaver, T. A. The evolution and explosion of massive stars. II. Explosive hydrodynamics and nucleosynthesis. *Astrophys. J. Suppl. Ser.* **101**, 181–235 (1995).
25. Rauscher, T., Heger, A., Hoffman, R. D. & Woosley, S. E. Nucleosynthesis in massive stars with improved nuclear and stellar physics. *Astrophys. J.* **576**, 323–348 (2002).
26. Limongi, M. & Chieffi, A. The nucleosynthesis of ^{26}Al and ^{60}Fe in solar metallicity stars extending in mass from 11 to 120 M_{\odot} : the hydrostatic and explosive contributions. *Astrophys. J.* **647**, 483–500 (2006).
27. Wanajo, S., Janka, H.-T. & Müller, B. Electron-capture supernovae as sources of ^{60}Fe . *Astrophys. J.* **774**, L6 (2013).
28. Doherty, C. L., Gil-Pons, P., Lau, H. H. B., Lattanzio, J. C. & Siess, L. Super and massive AGB stars. II. Nucleosynthesis and yields— $Z = 0.02, 0.008$ and 0.004 . *Mon. Not. R. Astron. Soc.* **437**, 195–214 (2014).
29. Frisch, P. C. *et al.* The galactic environment of the Sun: interstellar material inside and outside of the heliosphere. *Space Sci. Rev.* **146**, 235–273 (2009).
30. Farley, K. A., Vokrouhlický, D., Bottke, W. F. & Nesvorný, D. A late Miocene dust shower from the break-up of an asteroid in the main belt. *Nature* **439**, 295–297 (2006).
31. Bishop, S. *et al.* Search for supernova ^{60}Fe in the Earth's fossil record. *Am. Phys. Soc. April Meet.* **58**, X8.00002, <http://meetings.aps.org/link/BAPS.2013.APR.X8.2> (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by (1) the Austrian Science Fund (FWF), project number I428-N16 and within the ESF Eurogenesis programme; (2) the Australian Research Council (ARC), project number DP14100136; and (3) the Japan Society for the Promotion of Science (JSPS) KAKENHI grant number 26800161. J.F. acknowledges a stipend (Abschlusstipendium) from the University of Vienna. We thank the Antarctic Marine Geology Research Facility, Florida State University, USA (C. Sjunneskog) for providing the sediment cores, P. DeDecker (ANU) for help in selecting the cores; JOGMEC, Japan for supplying the crust; and P. Martínez Arbizu and M. Türkay for providing the nodules. Stable isotope measurements were performed by A. Ritter and S. Gurlit (HZDR) and V. Guilloat (CEREGE, France). We appreciate the support of M. Fröhlich, S. Akhmaliev, S. Pavetich, R. Ziegenrucker and P. Collon. We thank M. Lugaro and A. Karakas for information on (super)asymptotic-giant-branch stars and D. Bourlès on dating methods in deep-sea sediments. We thank D. Schumann for providing ^{60}Fe standard material.

Author Contributions A.W. initiated the study and wrote the main paper together with J.F., M.P. and L.K.F.; all authors were involved in the project and commented on the paper. A.W., with J.F., L.K.F. and S.R.W., organized the Eltanin sediment samples. N.K. and M.P. organized the crust samples. S.M. and U.L. organized the nodules. J.F. and S.M. were primarily responsible for sample preparation of the sediment and nodules and N.K. was responsible for the crusts. A.W., L.K.F. and S.G.T. performed the AMS measurements for ^{60}Fe at the ANU. P.S., S.R.W., J.F. and A.W. performed the ^{26}Al and ^{10}Be measurements at VERA. G.R., S.M. and J.F. performed ^{10}Be measurements at HZDR. N.K., M.H., H.M. and T.Y. performed ^{10}Be measurements at MALT. J.F., A.W. and N.K. performed the data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.W. (anton.wallner@anu.edu.au).

The locations of recent supernovae near the Sun from modelling ^{60}Fe transport

D. Breitschwerdt¹, J. Feige¹, M. M. Schulreich¹, M. A. de. Aveliz^{1,2}, C. Dettbarn³ & B. Fuchs³

The signature of ^{60}Fe in deep-sea crusts indicates that one or more supernovae exploded in the solar neighbourhood about 2.2 million years ago^{1–4}. Recent isotopic analysis is consistent with a core-collapse or electron-capture supernova that occurred 60 to 130 parsecs from the Sun⁵. Moreover, peculiarities in the cosmic ray spectrum point to a nearby supernova about two million years ago⁶. The Local Bubble of hot, diffuse plasma, in which the Solar System is embedded, originated from 14 to 20 supernovae within a moving group, whose surviving members are now in the Scorpius–Centaurus stellar association^{7,8}. Here we report calculations of the most probable trajectories and masses of the supernova progenitors, and hence their explosion times and sites. The ^{60}Fe signal arises from two supernovae at distances between 90 and 100 parsecs. The closest occurred 2.3 million years ago at present-day galactic coordinates $l = 327^\circ$, $b = 11^\circ$, and the second-closest exploded about 1.5 million years ago at $l = 343^\circ$, $b = 25^\circ$, with masses of 9.2 and 8.8 times the solar mass, respectively. The remaining supernovae, which formed the Local Bubble, contribute to a smaller extent because they happened at larger distances and longer ago (^{60}Fe has a half-life of 2.6 million years^{9,10}). There are uncertainties relating to the nucleosynthesis yields and the loss of ^{60}Fe during transport, but they do not influence the relative distribution of ^{60}Fe in the crust layers, and therefore our model reproduces the measured relative abundances very well.

It has been shown that supernovae from a moving stellar group, whose surviving members are now in the Sco–Cen association, were able to generate the Local Bubble^{7,11}. This hypothesis was further tested by performing three-dimensional high-resolution numerical simulations to follow the evolution of the Local Bubble from its formation until today and even into the future¹². It was found that the bubble's extension, as well as its ion column density ratios^{13,14}, matched ultraviolet observations (for example, FUSE¹⁵) remarkably well. To link the Local Bubble to the amount of ^{60}Fe discovered in a deep-sea ferromanganese crust², it was necessary to determine the distances to the supernovae. Therefore the masses of the supernova progenitors were calculated by exploiting the empirical fact that stars in clusters are likely to have a universal distribution in numbers over mass (initial mass function, IMF). We employed a variable-size binning to decrease the statistical bias in small samples by placing one star into each bin¹⁶, implying that 16 supernovae have exploded during the past 13 million years within the boundaries of the Local Bubble. Having derived the masses of the supernova progenitors, and assuming that all stars in the moving group were born at the same time, we were able to extract the age of the cluster⁷ from a comparison of the stars' positions in a Hertzsprung–Russell diagram with isochrones from stellar evolution calculations¹⁷. Since the main-sequence lifetime of a star is determined by its stellar mass for similar metallicities, as in a cluster, the explosion times are given by the difference of the cluster age and the respective lifespan of a supernova progenitor.

To pin down the explosion sites, we had to calculate the most probable trajectories of the group member stars, given the errors in the Hipparcos data¹⁸ and in the radial velocities. The uncertainties of stellar positions increase with distances, that is, when traced backwards in time from their current location in the Sco–Cen association. Since these uncertainties are measurement errors that are statistically independent, we can assume them to be Gaussian. This results in probability clouds (one for each explosion time) from which we extract the locations of highest probability and identify them as the explosion sites of the perished stars (see Fig. 1 for the two most recent supernovae). Thus we determined the present-day Galactic coordinates where the explosions occurred, with the two closest near $l = 327^\circ$, $b = 11^\circ$ and $l = 343^\circ$, $b = 25^\circ$ at distances of 90–100 pc (see Extended Data Table 1).

The transport of ^{60}Fe is strongly affected by the densities, velocities and pressures, both thermal and turbulent, of the ambient interstellar medium in general, and the Local Bubble in particular. Observationally, it was discovered from absorption in X-ray maps that the Local Bubble is interacting with the neighbouring Loop I bubble¹⁹, which restricts its expansion by counter-pressure. Hence it was necessary to produce a detailed account of the stellar content and the number of supernovae, as well as their explosion times and sites in Loop I. An analysis similar to that for the Local Bubble⁷ was carried out, but for a much larger volume of 800 pc in diameter. We found that two star clusters, Tr 10 and the association Vel OB2, have recently passed through the present volume of Loop I. Given a sample of 80 stars that might have entered this volume about 12.3 million years (Myr) ago, we estimated from an appropriate IMF²⁰ that, since then, 19 of the most massive stars should already have exploded. We derived the initial masses of those stars (and hence their explosion times and total ejecta mass) by assigning one star to each mass interval of the IMF as the statistically most probable distribution¹⁶. By the same token, the supernova explosion sites result from the trajectories of all the progenitor stars, which lie within the present Loop I.

To explain the deposition of ^{60}Fe on the ocean floor, we calculated both analytically and numerically the transport of ^{60}Fe in supernova blast waves from their sites of explosion to Earth. The Local Bubble shell evolution can be described analytically by a wind solution^{21,22}, with the density within the bubble increasing sharply proportional to $r^{9/2}$ or even more steeply (where r is the distance to the explosion site). We followed the expansion of shells in the interior of the Local Bubble, applying the Kahn approximation²³ (Fig. 2). Using a standard Sedov–Taylor^{24,25} solution for multiple explosions would give incorrect results, because the assumptions of negligible ambient pressure and constant density are no longer fulfilled. To estimate the abundance of ^{60}Fe in the crust, stellar yields of $(2\text{--}6) \times 10^{-5} M_\odot$ (where M_\odot is the solar mass) from nucleosynthesis models were used (see Methods). This range includes the yield proposed for electron-capture supernovae²⁶ of $3.6 \times 10^{-5} M_\odot$, which, together with their low progenitor masses, makes them prime candidates for the bulk of the ^{60}Fe deposition. The most likely mechanism of ^{60}Fe penetrating the heliosphere is its condensation into dust grains²⁷.

¹Department of Astronomy and Astrophysics, Berlin Institute of Technology, Hardenbergstraße 36, 10623 Berlin, Germany. ²Department of Mathematics, University of Évora, Rua Romão Ramalho 59, 7000 Évora, Portugal. ³Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12–14, 69120 Heidelberg, Germany.

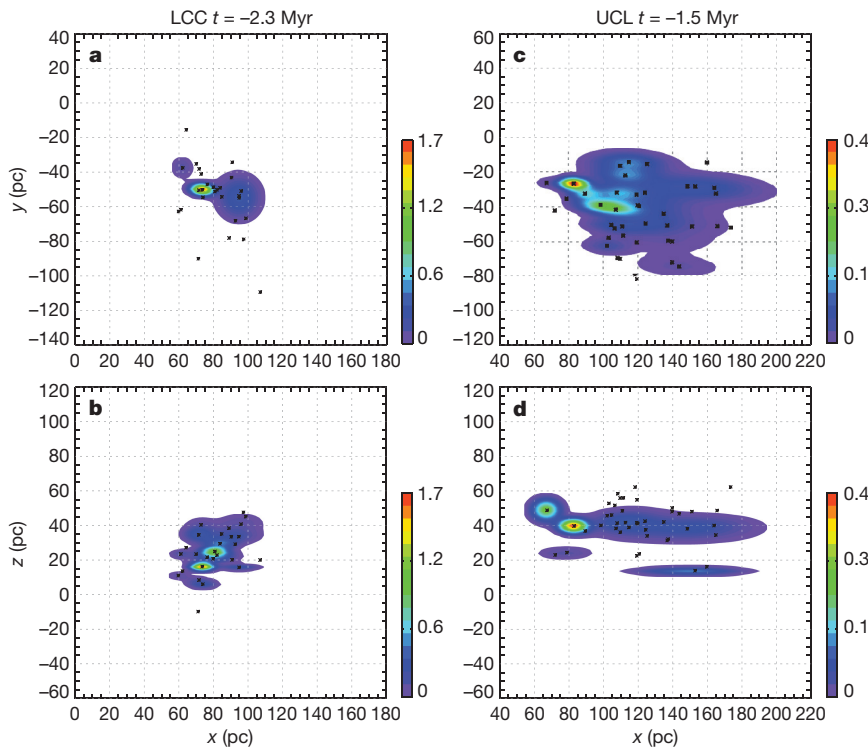


Figure 1 | Probability clouds of the two most recent supernovae. a–d, Clouds (not normalized, colour scale indicates the probability; the values of all voxels (volumetric pixels) add up to 1,000) in disk projection (a, c) and perpendicular (b, d) to it, include the location of highest probability (yellow–red colours on the scale). Still-existing stars of subgroups Upper Centaurus Lupus (UCL) and Lower Centaurus Crux (LCC) are shown as black points in projection. For $t = -2.3$ Myr (a, b) and $t = -1.5$ Myr (c, d) the sites of the perished stars are $(x, y, z) = (75, -49, 17)$ pc (distance of 91 pc from Earth) and $(x, y, z) = (83, -25, 41)$ pc (distance of 96 pc, see Extended Data Table 1), respectively.

The amount of dust, which forms in supernova remnants, is subject to potential losses during its transport to Earth. We have therefore adopted the value of $U \approx 0.6\%$ (ref. 2) for the ^{60}Fe survival fraction, which is similar to the value used by other authors⁵.

To determine the details of the contribution and the timescales of the ^{60}Fe deposition in the crust that is due to each individual supernova explosion, we performed three-dimensional high-resolution numerical hydrodynamics simulations with adaptive mesh refinement of the joint

evolution of the Local Bubble and Loop I. We allow the mass, momentum, energy and ^{60}Fe of each supernova to be directly injected into the numerical grid. As a background interstellar medium we have different homogeneous (down to sub-parsec resolution) and inhomogeneous turbulent supernova-driven interstellar environments, into which both superbubbles expand.

The dynamics of ^{60}Fe is followed via so-called passive scalars or tracers, which are quantities that behave like a drop of ink when dispersed

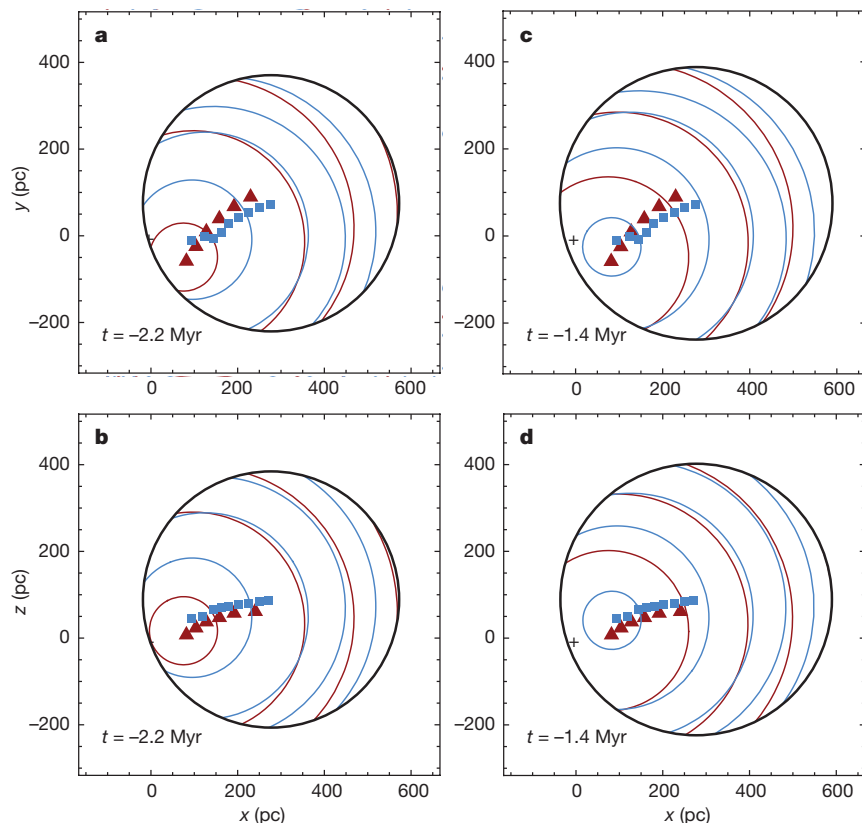


Figure 2 | Analytical model of ^{60}Fe -carrying shells at 1.4 Myr and 2.2 Myr before present. The outer black shell results from the Local Bubble evolution^{21,22}. The inner shells correspond to single remnants of stellar explosions (companion stars are now in UCL (blue) and LCC (red)), taking about 0.1 Myr to reach Earth. The blue squares and red triangles show the explosion sites of the UCL and LCC stars, respectively; the black cross denotes the position of the Sun. The layout of the panels corresponds to Fig. 1.

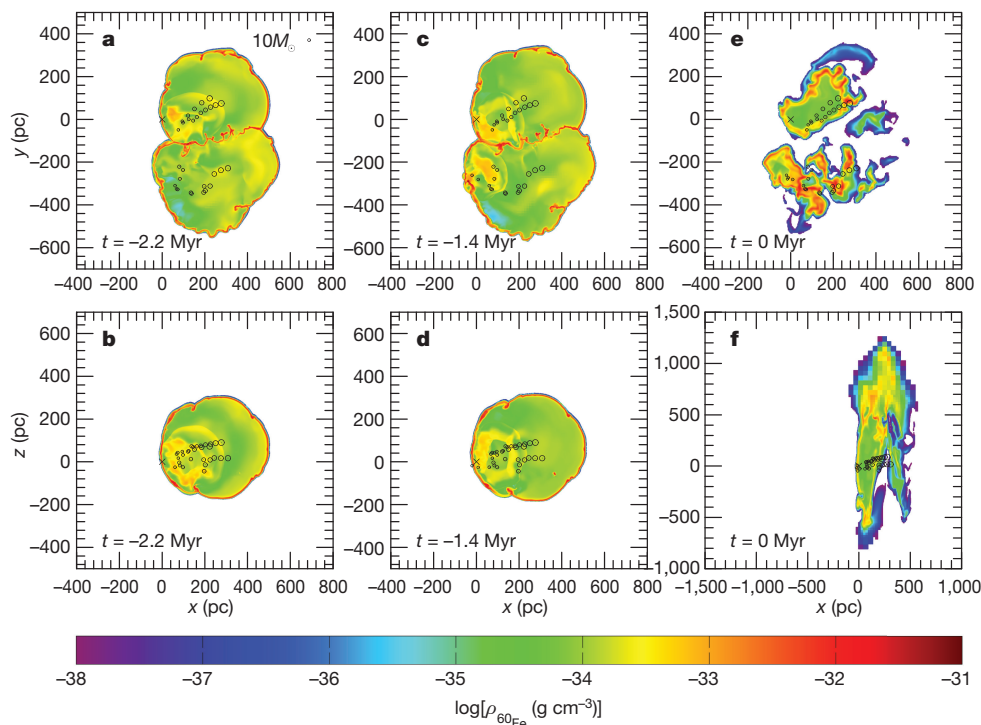


Figure 3 | Numerical simulations for ^{60}Fe distribution associated with the Local and Loop I superbubbles. **a–d**, The ^{60}Fe mass density distribution, $\rho_{^{60}\text{Fe}}$, for the bubbles expanding into a homogeneous ambient medium with a number density of $n_0 = 0.3 \text{ cm}^{-3}$ at $t = 2.2$ and 1.4 Myr before present. **e, f**, The same setup for an inhomogeneous background medium at $t = 0$ Myr. The supernova explosion centres are indicated by dotted circles (sizes corresponding to stellar masses), colour coding refers to logarithmic mass density of ^{60}Fe , and the cross to the location of the Solar System. The bubble interaction and reflected shocks are visible as reddish filaments.

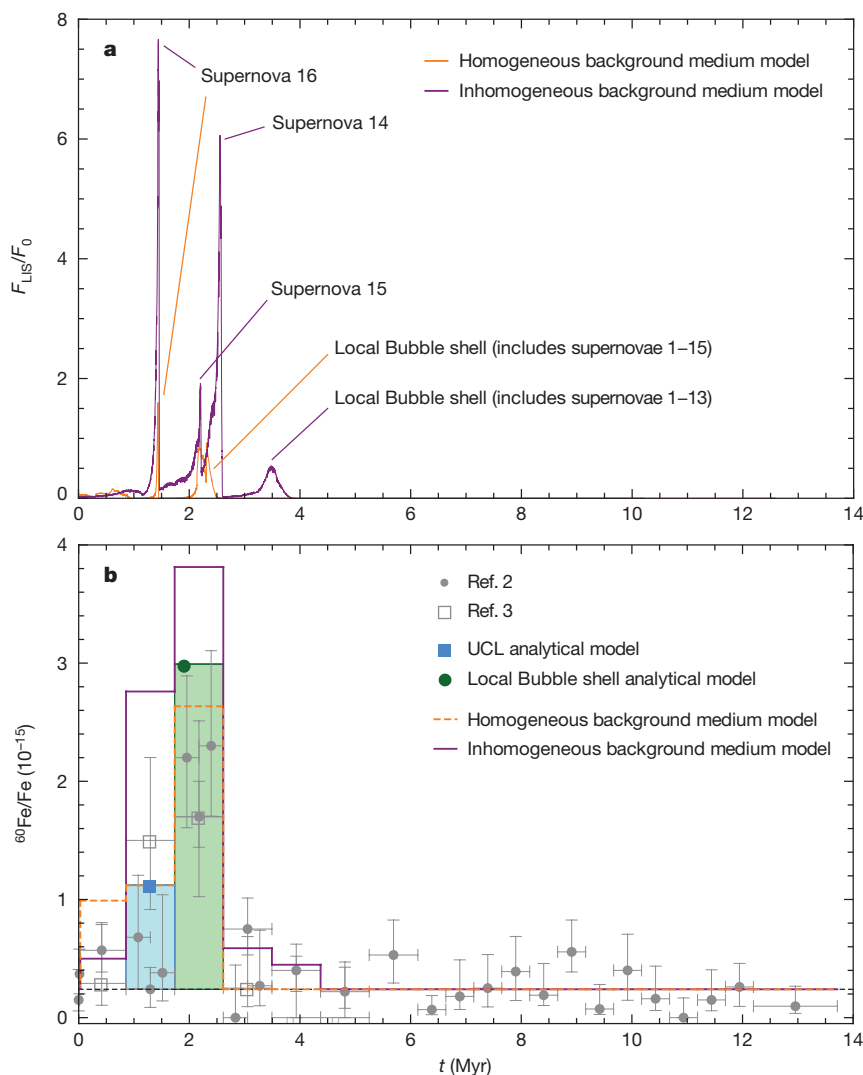


Figure 4 | Deposition of ^{60}Fe on Earth. **a**, Local interstellar fluence of ^{60}Fe atoms (F_{LIS} , normalized to $F_0 = 10^6 \text{ cm}^{-2}$) in the numerical model (homogeneous and inhomogeneous background medium). **b**, Comparison of $^{60}\text{Fe}/\text{Fe}$ measurements^{2,3}, to our best-fit models (vertical error bars are 1σ , horizontal bars correspond to time intervals, and dashed bottom line to instrumental background). The blue square and green dot indicate ^{60}Fe arrival times of shells; blue and green shaded histograms indicate the amount of $^{60}\text{Fe}/\text{Fe}$ deposited onto the crust (analytical model); numerical results are shown by orange and purple lines.

in a liquid. Figure 3 displays ^{60}Fe mass density distributions associated with the two superbubbles for two possible background media. By tagging the ^{60}Fe -enriched gas, we measured the flux of ^{60}Fe atoms at Earth's orbit over the whole simulation time, and then smeared it out into bins corresponding to the time resolution of the crust measurements.

Results for the homogeneous and inhomogeneous background media are presented in Fig. 4. Although in both cases the timing and intensity of the observed ^{60}Fe excess are well reproduced, the underlying physical processes are different. In the inhomogeneous case (solid purple line, Fig. 4b), the maximum signal is due to two individual supernova events (numbers 14 and 15 out of a total of 16; see Fig. 4a), which cross Earth's orbit twice as a result of shock reflection from the Local Bubble's outer shell. In the model with the homogeneous background medium (dashed orange line, Fig. 4b), it is the supershell of the Local Bubble itself that delivers the live ^{60}Fe content of all previous supernovae (numbers 1 to 15) at once (Fig. 4a), thereby stretching the deposition over several hundred thousand years.

Of all the simulations the model that fits the data best consists of an average ambient density of 0.3 cm^{-3} . The explosion sequence leads to the formation of the Local Bubble shell, transporting the highest ^{60}Fe content to Earth. The most recent explosion produces a much narrower but higher peak, resulting in a lower ^{60}Fe abundance at earlier times (Fig. 4). This is confirmed by our analytical studies with the same density, which show the distribution of ^{60}Fe -carrying shells for two specific times (Figs. 2 and 4b). Here, too, the Local Bubble shell reaches the Solar System at the time of the measured ^{60}Fe maximum, with a single supernova remnant shell following at later times. We find that the last two (three) supernovae contribute 47% (62%) to the amount of ^{60}Fe in the crust. Further evidence that the three latest explosions, between 1.5 Myr ago and 2.6 Myr ago (Extended Data Table 1), are electron-capture supernovae (with 3×10^{50} erg each) is given by recent analyses of the data obtained with the Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics (PAMELA; <http://pamela.roma2.infn.it>) satellite experiment. These show that the excess of cosmic-ray protons, antiprotons and positrons in the energy spectra points towards a total energy injection of 10^{51} erg about 2 Myr ago. These cosmic rays may also have produced an excess of long-lived radionuclides in meteoritic particles.

The absolute, but not relative, values of the ^{60}Fe distribution in the crust layers depend mainly on two factors, which are beyond our calculations, and serve as input parameters: (1) the ^{60}Fe yields of supernovae, which are mass dependent and taken from stellar evolution calculations, and (2) the transport factors, which describe the deposition of ^{60}Fe in the crust. Once better values become available, our simulations can be further constrained and refined. The calculations reported here show that it is possible to form a coherent picture of the evolution of the Local Bubble and the ^{60}Fe deposition in the crust that is consistent with observations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 July 2015; accepted 10 February 2016.

1. Knie, K. *et al.* Indication for supernova produced ^{60}Fe activity on Earth. *Phys. Rev. Lett.* **83**, 18–21 (1999).
2. Knie, K. *et al.* ^{60}Fe anomaly in a deep-sea manganese crust and implications for a nearby supernova source. *Phys. Rev. Lett.* **93**, 171103 (2004).
3. Fitoussi, C. *et al.* Search for supernova-produced ^{60}Fe in a marine sediment. *Phys. Rev. Lett.* **101**, 121101 (2008).
4. Bishop, S. *et al.* Search for supernova ^{60}Fe in the Earth's fossil record. In *American Physical Society April Meeting 2013* **58**, abstr. X8.00002, <http://meetings.aps.org/link/BAPS.2013.APR.X8.2> (2013).
5. Fry, B. J., Fields, B. D. & Ellis, J. R. Astrophysical shrapnel: discriminating among near-Earth stellar explosion sources of live radioactive isotopes. *Astrophys. J.* **800**, 71 (2015).

6. Kachelrieß, M., Neronov, A. & Semikoz, D. V. Signatures of a two million year old supernova in the spectra of cosmic ray protons, antiprotons, and positrons. *Phys. Rev. Lett.* **115**, 181103 (2015).
7. Fuchs, B., Breitschwerdt, D., de Avillez, M. A., Dettbarn, C. & Flynn, C. The search for the origin of the Local Bubble redivivus. *Mon. Not. R. Astron. Soc.* **373**, 993–1003 (2006).
8. Benítez, N., Maíz-Apellániz, J. & Canelles, M. Evidence for nearby supernova explosions. *Phys. Rev. Lett.* **88**, 081101 (2002).
9. Rugel, G. *et al.* New measurement of the ^{60}Fe half-life. *Phys. Rev. Lett.* **103**, 072502 (2009).
10. Wallner, A. *et al.* Settling the half-life of ^{60}Fe : fundamental for a versatile astrophysical chronometer. *Phys. Rev. Lett.* **114**, 041101 (2015).
11. Berghöfer, T. W. & Breitschwerdt, D. The origin of the young stellar population in the solar neighborhood—a link to the formation of the Local Bubble? *Astron. Astrophys.* **390**, 299–306 (2002).
12. Breitschwerdt, D. & de Avillez, M. A. The history and future of the Local and Loop I bubbles. *Astron. Astrophys.* **452**, L1–L5 (2006).
13. de Avillez, M. A. & Breitschwerdt, D. The distribution of Li-like ions in the Local Bubble. *Astrophys. J.* **697**, L158–L161 (2009).
14. de Avillez, M. A. & Breitschwerdt, D. Non-equilibrium ionization modeling of the Local Bubble. I. Tracing C IV, N V, and O VI ions. *Astron. Astrophys.* **539**, L1 (2012).
15. Savage, B. D. & Lehnner, N. Properties of O VI absorption in the local interstellar medium. *Astrophys. J.* **162** (Suppl.), 134–160 (2006).
16. Maíz-Apellániz, J. & Úbeda, L. Numerical biases on initial mass function determinations created by binning. *Astrophys. J.* **629**, 873–880 (2005).
17. Schaller, G., Schaerer, D., Meynet, G. & Maeder, A. New grids of stellar models from 0.8 to 120 M_{\odot} at $Z = 0.020$ and $Z = 0.001$. *Astron. Astrophys. Suppl. Ser.* **96**, 269–331 (1992).
18. European Space Agency. *The HIPPARCOS and TYCHO Catalogues. Astrometric and Photometric Star Catalogues derived from the ESA HIPPARCOS Space Astrometry Mission.* ESA Special Publication SP-1200, http://www.rssd.esa.int/SA/HIPPARCOS/docs/vol1_all.pdf (ESA, 1997).
19. Egger, R. J. & Aschenbach, B. Interaction of the Loop I supershell with the Local Hot Bubble. *Astron. Astrophys.* **294**, L25–L28 (1995).
20. Massey, P., Johnson, K. E. & DeGioia-Eastwood, K. The initial mass function and massive star evolution in the OB associations of the northern Milky Way. *Astrophys. J.* **454**, 151–171 (1995).
21. Weaver, R., McCray, R., Castor, J., Shapiro, P. & Moore, R. Interstellar bubbles. II—structure and evolution. *Astrophys. J.* **218**, 377–395 (1977).
22. McCray, R. & Kafatos, M. Supershells and propagating star formation. *Astrophys. J.* **317**, 190–196 (1987).
23. Kahn, F. D. in *IAU Colloq. 166: The Local Bubble and Beyond* (eds Breitschwerdt, D., Freyberg, M. J. & Truemper, J.) Vol. 506 of Lecture Notes in Physics 483–494 (Springer, 1998).
24. Sedov, L. I. *Similarity and Dimensional Methods in Mechanics* 10th edn, 242–251 (CRC Press, 1993).
25. Taylor, G. The formation of a blast wave by a very intense explosion. II. The atomic explosion of 1945. *Proc. R. Soc. Lond. Ser. A* **201**, 175–186 (1950).
26. Wanajo, S., Janka, H.-T. & Müller, B. Electron-capture supernovae as sources of ^{60}Fe . *Astrophys. J.* **774**, L6 (2013).
27. Athanassiadou, T. & Fields, B. D. Penetration of nearby supernova dust in the inner solar system. *New Astron.* **16**, 229–241 (2011).

Acknowledgements D.B., M.A.de.A. and M.M.S. acknowledge funding by the DFG priority program 1573 “Physics of the Interstellar Medium”. We thank U. Bolick for help during the preparation of the manuscript and R. Teyssier for discussions on details of the RAMSES code.

Author Contributions D.B. worked out the model, and led the research and the paper writing. J.F. carried out analytic calculations, interpreted the crust data, calculated the IMF and produced Figs. 1 and 2 and the lower part of Fig. 4, and Extended Data Figs 1–5. M.M.S. performed extensive numerical simulations on the background interstellar medium, the evolution of the Local Bubble and Loop I bubbles and the ^{60}Fe transport and produced Fig. 3 and the upper part of Fig. 4. M.A.de.A. was involved in the interpretation of the data and the numerical simulations. C.D. carried out the analysis of the moving group stars, calculated the trajectories of both Local Bubble and Loop I progenitor stars and wrote a program to determine the probability distributions. B.F. worked analytically on the epicyclic equations and carried out the cumulative distribution function calculations for the most probable trajectories.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (breitschwerdt@astro.physik.tu-berlin.de).

METHODS

Analytic modelling. In the following we focus on the supernovae closest to Earth. Of course, on a considerably larger scale than 100 pc, more recent (<300,000 yr ago) explosions have occurred, such as Vela and Geminga, which may have left signatures in ^{14}C , cosmic rays and Earth's biosphere^{28–31}, but not in ^{60}Fe owing to the large distances from Earth.

In case of the nearby stellar moving group, we closely examined each star's velocity, including its uncertainty, and calculated the most probable path for the group inside the Local Bubble. Since the progenitors could not be observed directly, we assume them to have followed this trajectory together with the other group members. The putative explosion sites were calculated for each star (see Fig. 1 and Extended Data Table 1). Coordinates and velocities from still-existing stars of subgroups UCL and LCC were inserted into the epicyclic equations of motion^{32,33} to trace their trajectories back in time⁷. The cumulative distribution function can be written as

$$F = F_{\text{plan}} F_{\text{vert}} \quad (1)$$

where we assume the planar and vertical stellar motions to be statistically independent and therefore decoupled. Hence, the components of the distribution function are described by

$$F_{\text{plan}} = \frac{1}{(2\pi)^{3/2} \sigma_x \sigma_u \sigma_v} \int_{-\infty}^{\infty} dx_0 \int_{-\infty}^{\infty} du_0 \int_{-\infty}^{\infty} dv_0 p(x, y | x_0, y_0, u_0, v_0) \times \exp \left[-\frac{1}{2} \left(\frac{(x_0 - x_{\text{obs}})^2}{\sigma_x^2} + \frac{(u_0 - u_{\text{obs}})^2}{\sigma_u^2} + \frac{(v_0 - v_{\text{obs}})^2}{\sigma_v^2} \right) \right] \quad (2)$$

and

$$F_{\text{vert}} = \frac{1}{2\pi \sigma_z \sigma_w} \int_{-\infty}^{\infty} dz_0 \int_{-\infty}^{\infty} dw_0 p(z | w_0, z_0) \times \exp \left[-\frac{1}{2} \left(\frac{(z_0 - z_{\text{obs}})^2}{\sigma_z^2} + \frac{(w_0 - w_{\text{obs}})^2}{\sigma_w^2} \right) \right] \quad (3)$$

where x_{obs} , z_{obs} , u_{obs} , v_{obs} , and w_{obs} are the observed quantities of the positional and velocity components⁷. Here x , y and z denote the direction towards the Galactic Centre, the Galactic rotation and the North Galactic Pole, respectively. The associated σ_i are the corresponding Gaussian uncertainties. The positional error of the y -component, σ_y , is small compared to the error of the x -component, that is, $y_0 = y_{\text{obs}}$, and is therefore neglected. The conditional probability p is given by a δ distribution

$$p(x, y | x_0, y_0, u_0, v_0) = \delta(x - x_t) \delta(y - y_t) \quad (4)$$

$$p(z | w_0, z_0) = \delta(z - z_t)$$

The index t denotes the components of stellar motion at time t calculated by the epicyclic equations^{32,33}.

The observed uncertainties increase with distance, that is, when traced backwards in time. This results in probability clouds for each explosion time, from which a location of highest probability is extracted and assumed to be the explosion site of the perished star. Figure 1 depicts such clouds for the two most recent supernovae.

The radius R_{LB} of the outer Local Bubble shell at time t after the explosion of the first supernova was calculated by²¹

$$R_{\text{LB}} = 132 \left(\frac{N_* E_{51}}{n_0} \right)^{1/5} t_7^{3/5} \text{ parsecs} \quad (5)$$

Here, N_* is the number of supernovae, E_{51} is the explosion energy in units of 10^{51} erg, n_0 is the outer interstellar gas number density in units of cm^{-3} , into which the shell is expanding, and t_7 is the evolution time in units of 10^7 yr. For densities higher than 0.3 cm^{-3} , the shell is not able to reach the Solar System because, in the analytical model, the first supernova had to be assumed to explode in the Local Bubble centre. The same also happens in the numerical simulation (although no restriction about the first explosion site had to be made) of the Local Bubble expansion into a homogeneous medium; hence this interstellar gas density is chosen as an upper limit. The free expansion of an individual shell, calculated in the analytical model, occurred into a low-density Local Bubble medium^{7,11} of $5 \times 10^{-3} \text{ cm}^{-3}$. This phase lasts approximately 6,000 yr and covers a distance of about 25 pc, with variations depending on the ejected mass. The subsequent energy-driven expansion is calculated by an approximation due to Kahn²³, taking into account the density gradient and the counter-pressure of the ambient bubble medium

$$R_{\text{Kahn}} = \left(\frac{(n+5)(2n+7) E_{\text{SN}}}{6\pi\Omega} \right)^{\frac{1}{n+5}} t^{\frac{2}{n+5}} \quad (6)$$

where the density profile of the Local Bubble medium is $\rho = \Omega r^n$ with $n = 9/2$ (for a single supernova shell) and E_{SN} is the supernova explosion energy. The time-dependent expression for Ω includes the density n_0 of the ambient medium and the beginning of the energy-driven supernova remnant evolution. When written in the dimensionless units introduced above, equation (6) reads as

$$R_{\text{Kahn}} = 448 \left(\frac{E_{51}}{n_0} \right)^{2/19} t_7^{0.478} \text{ parsecs} \quad (7)$$

To obtain the correct hierarchy of shells, it is required that the radius R_{SNR} of each supernova remnant obeys at all times and positions

$$R_{\text{SNR}} = \min(R_{\text{LB}}, R_{\text{Kahn}}) \quad (8)$$

indicating that equation (7) only holds until an interior supernova shell hits the outer Local Bubble shell.

The amount of ejected ^{60}Fe was estimated using several nucleosynthesis models^{34–37}. Depending on the input (such as cross-sections, mass loss and rotation) into these models, the ^{60}Fe yield is scattered over a wide range between $10^{-6} M_{\odot}$ and $10^{-4} M_{\odot}$. The mass-dependent yields were plotted and fitted by an exponential function. Thus, for each initial mass of an exploded star a yield of the order of 10^{-5} was obtained (see Extended Data Table 1 and Extended Data Fig. 1).

The ^{60}Fe fluence, which is the number of ^{60}Fe atoms per cm^2 that is incorporated into the ferromanganese crust, is calculated by³⁸

$$F = \frac{U}{4} \frac{M_{\text{ej}}}{4\pi A m_p r^2} e^{-t \ln(2)/t_{1/2}} \quad (9)$$

where r is the distance of the supernova explosion to the Solar System, M_{ej} is the ejected ^{60}Fe mass, A its atomic mass number, m_p the proton mass, and U the uptake factor. U is divided by a factor of 4 to account for the ratio of Earth's cross-section to its surface area. The formula includes the exponential decay of ^{60}Fe atoms with time t , where $t_{1/2}$ is its half-life. The fluence is converted into $^{60}\text{Fe}/\text{Fe}$ ratios by normalizing it to the amount of stable Fe in the crust of 15.27 wt% (ref. 39).

Finally, only a fraction of ^{60}Fe , the so-called uptake factor U , is eventually incorporated into the crust. An earlier proposed value of U was about 0.6% (ref. 2), but more recent analyses point to a much higher value⁴⁰. In our models we have adopted the original value, which takes into account not only the uptake into the crust, but further losses during transport and the yield uncertainties. It may therefore be interpreted as the ^{60}Fe survival fraction, and is close to the value of 0.5% (ref. 5).

Setup of the numerical model. The numerical simulations have been performed with a modified version of the massively parallel multi-purpose astrophysical fluid code RAMSES⁴¹. It uses a second-order unsplit Godunov scheme for solving the hydrodynamical equations on an adaptive octree grid, and a Particle-Mesh solver for computing the trajectories of collisionless particles, which represent individual massive stars in our setup (see below). The computational domain is cubic with a side length of 3 kpc. In the model with the homogeneous background medium, all boundaries are treated as periodic, whereas in the model with the inhomogeneous background medium periodic and outflow boundaries are applied at the vertical and horizontal sides, respectively. The inhomogeneous background medium is self-consistently generated by exposing an initial interstellar gas distribution, derived from observations⁴², for 180 Myr (a timescale comparable to that required to reach a dynamical equilibrium⁴³) to the combined effects of the Galactic gravitational field^{44,45}, heating by the interstellar radiation field and cooling by collisional ionization equilibrium processes (all modelled using the spectral synthesis code CLOUDY⁴⁶), as well as winds of massive stars, and supernovae.

Our simulations currently neglect magnetic fields, because their dynamical effects are, when compared to turbulence, presumably small on the scale of the dynamical evolution of the isolated patch of interstellar matter⁴³, and in particular the Local Bubble. In the case of the Local Bubble, the magnetic field should be concentrated in the supershell, where it has been swept, together with the matter, by successive supernova explosions. The Local Bubble interior should thus be almost field-free. Stars in the mass range considered of $0.5 \leq M/M_{\odot} \leq 150$ are assumed to form at a Galactic rate⁴⁷ when the gas in a computational cell exceeds (falls below) a certain density (temperature) threshold, with the initial stellar mass spectrum obeying an IMF⁴⁸. However, only candidates for Type II supernovae ($10 \leq M/M_{\odot} \leq 30$) are converted into actual star particles, which is certainly justified for the young moving groups, which generated the Local Bubble and Loop I. Both their

(main-sequence) lifetimes and their mass loss (due to stellar wind and the supernova) are taken to be functions of their initial masses³⁷. The stellar wind velocity is assumed to be constant and is set equal to the canonical value of $2,000 \text{ km s}^{-1}$. Freshly spawned stellar particles receive in addition a random drift velocity⁴⁹ of 5 km s^{-1} , to account for the observed velocity dispersion. Owing to the high computational demands of such a complex simulation, the finest numerical grid is here somewhat coarser (2.9 pc) than in the model with the homogeneous background medium. As was tested by means of calculating filling factors and probability distribution functions, the inhomogeneous medium, however, serves as a much more realistic environment for studying the Local Bubble evolution scenario.

Using a model for non-self-gravitating compressible turbulence⁵⁰, we analysed the properties of turbulence in our simulated patch of the Galactic disk. We also⁵¹ found that turbulence is injected at the scales of breaking open superbubbles, and, furthermore, that the structure functions up to higher orders are nicely matched by a scaling law specifically derived for supersonic turbulence, in which shocks represent the most dissipative structures⁵².

As noted before, ^{60}Fe is treated as a passive scalar, which is justified by its low concentration, precluding any relevant back-reactions on the flow (for example, buoyancy). The spatiotemporal evolution of the ^{60}Fe distribution can therefore be approximately described by an advection–diffusion equation of the form⁵³

$$\frac{\partial Z}{\partial t} + (\mathbf{u} \cdot \nabla)Z = \alpha \nabla^2 Z \quad (10)$$

where Z is the ^{60}Fe mass fraction, \mathbf{u} is the fluid velocity, and α is the diffusivity of the contaminant (which is assumed here to be isotropic). Besides taking into account the radioactive decay of ^{60}Fe (using its latest derived half-life^{9,10}), we have actively modelled only the left-hand side of this equation, which is motivated by the fact that, particularly in the interstellar medium, diffusive effects are usually restricted to the very small scale (the so-called microscale) of turbulence. However, our simulations still feature non-zero diffusion of ^{60}Fe arising from the numerical scheme itself. Since such numerical diffusion has been demonstrated to operate in general faster and on larger scales than its physical counterpart⁵⁴, the timescale of mixing in our simulations represents a lower limit to the mixing timescale resulting from physical diffusion.

Code availability. The RAMSES code is available for free download online (<https://bitbucket.org/rteyssi/ramses>). The newly written or modified routines of the code, as well as the analytical model, which was programmed in MATHEMATICA, that are required for producing the specific setup and results discussed here, are still being improved and extended, and will therefore be released in the future.

Modelling ^{26}Al and ^{53}Mn signals in the ferromanganese crust. ^{60}Fe is the most suitable long-lived radionuclide for detecting recent nearby supernova explosions because (1) it is not produced by terrestrial sources, (2) its extraterrestrial influx via interstellar dust particles and micrometeorites is negligible, and (3) all primordial ^{60}Fe has decayed since the formation of the Solar System.

Here we present the distributions of two other candidate isotopes, ^{26}Al ($t_{1/2} = 0.7 \text{ Myr}$) and ^{53}Mn ($t_{1/2} = 3.7 \text{ Myr}$), in the ferromanganese crust. Again, we performed analytical calculations using the supernova remnant expansion model described by Kahn²³, and assume that the outer boundary evolves as a superbubble shell²¹. Therefore, stellar explosion times and arrival times are identical to the ^{60}Fe model. In addition, the survival fraction of 0.6% was adopted from this model. Here, to be conservative, we used the highest possible yields for ^{26}Al and ^{53}Mn (see Extended Data Figs 2 and 3)^{35,36}.

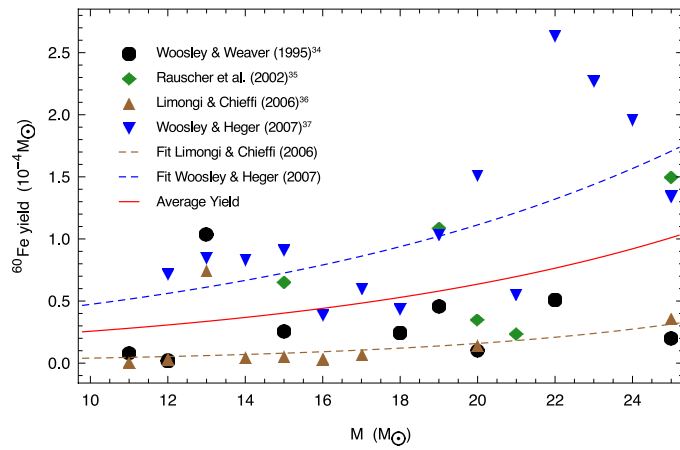
Furthermore, relative supernova abundances of $^{60}\text{Fe}/^{26}\text{Al}$ are suggested⁵⁵ to lie between 0.6 and 23. As a second approach, we have used the upper and lower limits of these ratios to scale the measured ^{60}Fe signal in the 1.7–2.6-Myr-old layer of the ferromanganese crust and estimate the expected ^{26}Al abundance, taking into account the different half-lives.

In this case it is important to consider the large atmospheric production of ^{26}Al , which arises from spallation of mainly Ar atoms by cosmic rays in Earth's atmosphere. The deposition of ^{26}Al onto the layers is displayed in Extended Data Fig. 4. The concentration of the radionuclide is normalized to the stable Al content of 0.6 wt% in the crust (K. Knie, private communication, November 2015). The terrestrial ^{26}Al background, usually of the order of $^{26}\text{Al}/^{27}\text{Al} \approx 10^{-13}$ at the surface^{56,57}, overwhelms any possible supernova signal in all scenarios considered here. Clearly, a detection of a ^{26}Al supernova peak in the crust is challenging, since the accelerator mass spectrometry measurement uncertainties are larger than the modelled supernova signature. The ^{26}Al estimate presented here is consistent with the ^{26}Al value measured in the ferromanganese crust (K. Knie, private communication, November 2015) as well as with recent measurements performed in deep-sea sediments, where no supernova signal was found⁵⁸.

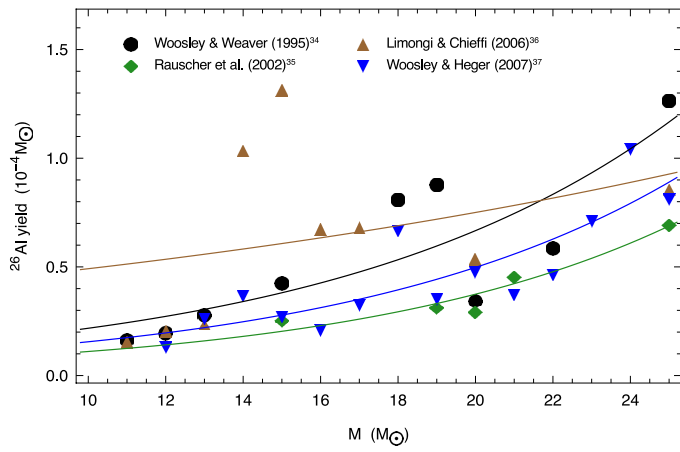
The situation is similar for ^{53}Mn . Extraterrestrial interstellar dust particles and micrometeoritic influx may distribute ^{53}Mn on Earth and hide a potential supernova

signal (Extended Data Fig. 5). The ^{53}Mn concentration was normalized to the stable ^{55}Mn content of 22.67 wt% and the results are in agreement with ^{53}Mn accelerator mass spectrometry measurements performed in the ferromanganese crust³⁹. Again, the measurement error bars are larger than the calculated supernova signal and therefore the extraterrestrial background hampers the detection of supernova-produced ^{53}Mn .

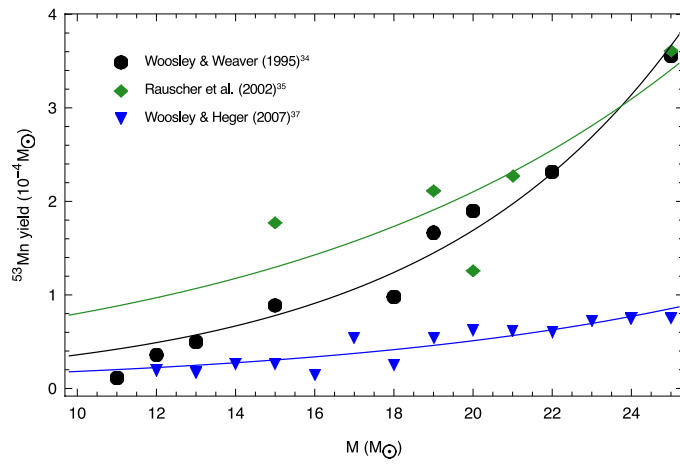
28. Firestone, R. B. Observation of 23 supernovae that exploded <300 pc from Earth during the past 300 kyr. *Astrophys. J.* **789**, 29 (2014).
29. Erlykin, A. D. & Wolfendale, A. W. Cosmic ray antiprotons and the single source model. *J. Phys. G* **42**, 115202 (2015).
30. Melott, A. L., Usoskin, I. G., Kovaltsov, G. A. & Laird, C. M. Has the Earth been exposed to numerous supernovae within the last 300 kyr? *Int. J. Astrobiol.* **14**, 375–378 (2015).
31. Melott, A. L. A possible role for stochastic radiation events in the systematic disparity between molecular and fossil dates. Preprint at <http://arxiv.org/abs/1505.08125> (2015).
32. Lindblad, B. Galactic dynamics. *Handb. Phys.* **11/53**, 21–99 (1959).
33. Wielen, R. in *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology* (eds Schaifers, K. & Voigt, H. H.) Group VI, Vol. 2, 225–227 (Springer, 1982).
34. Woosley, S. E. & Weaver, T. A. The evolution and explosion of massive stars. II. Explosive hydrodynamics and nucleosynthesis. *Astrophys. J.* **101** (Suppl.), 181 (1995).
35. Rauscher, T., Heger, A., Hoffman, R. D. & Woosley, S. E. Nucleosynthesis in massive stars with improved nuclear and stellar physics. *Astrophys. J.* **576**, 323–348 (2002).
36. Limongi, M. & Chieffi, A. The nucleosynthesis of ^{26}Al and ^{60}Fe in solar metallicity stars extending in mass from 11 to $120 M_{\odot}$: the hydrostatic and explosive contributions. *Astrophys. J.* **647**, 483–500 (2006).
37. Woosley, S. E. & Heger, A. Nucleosynthesis and remnants in massive stars of solar metallicity. *Phys. Rep.* **442**, 269–283 (2007).
38. Fields, B. D., Hochmuth, K. A. & Ellis, J. Deep-ocean crusts as telescopes: using live radioisotopes to probe supernova nucleosynthesis. *Astrophys. J.* **621**, 902–907 (2005).
39. Poutvtsev, M. *Extraterrestrisches ^{53}Mn in hydrogenetischen Mangankrusten*. PhD thesis, Technische Univ. München, <https://mediatum.ub.tum.de/?id=610694> (2007).
40. Feige, J. et al. AMS measurements of cosmogenic and supernova-ejected radionuclides in deep-sea sediment cores. In *Eur. Phys. J. Web Conf.* **63**, 03003, <http://dx.doi.org/10.1051/epjconf/20136303003> (2013).
41. Teyssier, R. Cosmological hydrodynamics with adaptive mesh refinement. *Astron. Astrophys.* **385**, 337–364 (2002).
42. Ferrière, K. Global model of the interstellar medium in our galaxy with new constraints on the hot gas component. *Astrophys. J.* **497**, 759–776 (1998).
43. de Avillez, M. A. & Breitschwerdt, D. Volume filling factors of the ISM phases in star forming galaxies. I. The role of the disk-halo interaction. *Astron. Astrophys.* **425**, 899–911 (2004).
44. Kuijken, K. & Gilmore, G. The mass distribution in the Galactic Disc. I—A technique to determine the integral surface mass density of the disc near the sun. *Mon. Not. R. Astron. Soc.* **239**, 571–603 (1989).
45. Kuijken, K. & Gilmore, G. The mass distribution in the Galactic Disc. II—Determination of the surface mass density of the Galactic Disc near the Sun. *Mon. Not. R. Astron. Soc.* **239**, 605–649 (1989).
46. Ferland, G. J. et al. CLOUDY 90: numerical simulation of plasmas and their spectra. *Publ. Astron. Soc. Pac.* **110**, 761–778 (1998).
47. Robitaille, T. P. & Whitney, B. A. The present-day star formation rate of the Milky Way determined from Spitzer-detected young stellar objects. *Astrophys. J.* **710**, L11–L15 (2010).
48. Salpeter, E. E. The luminosity function and stellar evolution. *Astrophys. J.* **121**, 161–167 (1955).
49. Blaauw, A. The O associations in the solar neighborhood. *Astron. Astrophys. Rev.* **2**, 213–246 (1964).
50. Fleck, R. C. Scaling relations for the turbulent, non-self-gravitating, neutral component of the interstellar medium. *Astrophys. J.* **458**, 739 (1996).
51. de Avillez, M. A. & Breitschwerdt, D. The generation and dissipation of interstellar turbulence: results from large-scale high-resolution simulations. *Astrophys. J.* **665**, L35–L38 (2007).
52. Boldyrev, S. Kolmogorov-Burgers model for star-forming turbulence. *Astrophys. J.* **569**, 841–845 (2002).
53. Davidson, P. A. *Turbulence: An Introduction for Scientists and Engineers* 1st edn, 234–235 (Oxford Univ. Press, 2004).
54. de Avillez, M. A. & Mac Low, M.-M. Mixing timescales in a supernova-driven interstellar medium. *Astrophys. J.* **581**, 1047–1060 (2002).
55. Wasserburg, G. J., Gallino, R. & Busso, M. A test of the supernova trigger hypothesis with ^{60}Fe and ^{26}Al . *Astrophys. J.* **500**, L189–L193 (1998).
56. Wang, L., Ku, T. L., Luo, S., Southon, J. R. & Kusakabe, M. ^{26}Al - ^{10}Be systematics in deep-sea sediments. *Geochim. Cosmochim. Acta* **60**, 109–119 (1996).
57. Sharma, P., Klein, J., Middleton, R. & Church, T. M. ^{26}Al and ^{10}Be in authigenic marine minerals. *Nucl. Instrum. Methods Phys. Res. B* **29**, 335–340 (1987).
58. Feige, J. *Supernova-produced Radionuclides in Deep-sea Sediments Measured with AMS*. PhD thesis, Univ. Vienna, <http://othes.univie.ac.at/35089> (2014).



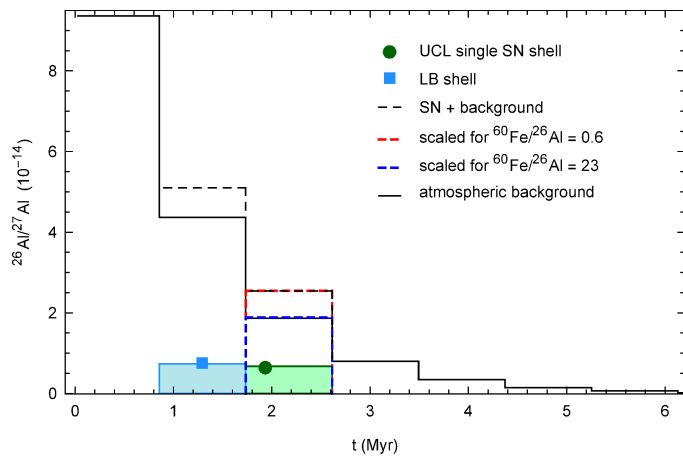
Extended Data Figure 1 | ^{60}Fe yields from various nucleosynthesis models^{34–37}. Given that the yields are mass-dependent, an average between the highest³⁷ and lowest³⁶ was calculated and extrapolated towards the lower mass range. Superscripts in the legend refer to reference numbers.



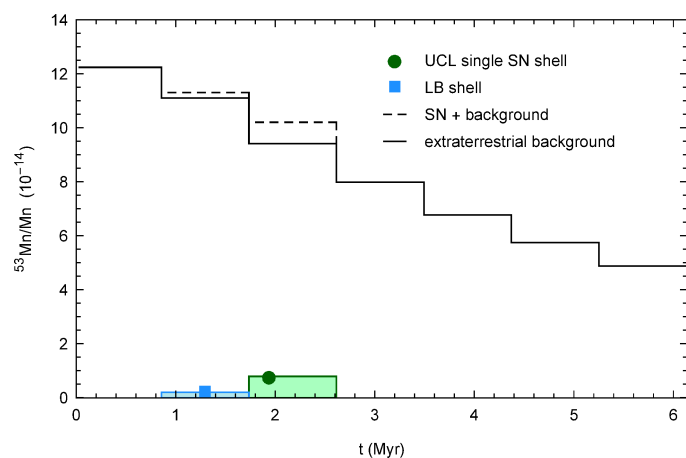
Extended Data Figure 2 | ^{26}Al yields from various nucleosynthesis models^{34–37}. To be conservative, the highest yields³⁶ were fitted and extrapolated towards the lower mass range.



Extended Data Figure 3 | ^{53}Mn yields from various nucleosynthesis models^{34,35,37}. Again, to be conservative, the highest yields³⁵ were fitted and extrapolated towards the lower mass range.



Extended Data Figure 4 | ^{26}Al in the ferromanganese crust. The largest fraction arises from atmospheric production, hiding a possible supernova signal. The arrival times of the supernova shells (analytical model with input parameters identical to the ^{60}Fe model) are indicated by a blue square (UCL) and a green dot (LB, Local Bubble shell), whereas the blue and green shaded histograms represent the amount of ^{26}Al deposited onto the crust, normalized to ^{27}Al . The resulting supernova signal (SN) is added to the atmospheric background (black dashed line). The measured ^{60}Fe concentration² in the layer showing the largest peak (red (blue) dashed lines) was used to scale towards the ^{26}Al content using supernova $^{60}\text{Fe}/^{26}\text{Al}$ ratios⁵⁵ of 0.6 (23). Here, the atmospheric background has been added to the signal.



Extended Data Figure 5 | ^{53}Mn in the ferromanganese crust. The largest fraction is produced from extraterrestrial dust and micrometeorite influx, hiding a possible supernova signal. Again, the analytical model with input parameters identical to the ^{60}Fe model was used. The notation is the same as in Fig. 4.

Extended Data Table 1 | Supernova explosions that created the Local Bubble

t_{SN}	m (M_{\odot})	M_{ej} ($10^{-5} M_{\odot}$)	x (pc)	y (pc)	z (pc)	D (pc)	l ($^{\circ}$)	b ($^{\circ}$)	α	δ	sc
-12.6 ²	19.86	6.3	277	75	89	300	15.15	17.23	17 ^h 17 ^m	-7 $^{\circ}$ 09 ^m	Oph
-12.0 ³	18.61	5.5	223	99	71	254	23.94	16.22	17 ^h 37 ^m	-0 $^{\circ}$ 21 ^m	Oph
-11.3 ²	17.34	5.0	251	67	87	274	14.95	18.52	17 ^h 12 ^m	-6 $^{\circ}$ 39 ^m	Oph
-10.0 ²	15.41	4.2	227	57	83	248	14.10	19.53	17 ^h 07 ^m	-6 $^{\circ}$ 48 ^m	Oph
-10.0 ³	15.36	4.1	185	77	67	211	22.60	18.49	17 ^h 27 ^m	-0 $^{\circ}$ 23 ^m	Oph
-8.7 ²	13.89	3.6	203	45	79	222	12.50	20.80	17 ^h 00 ^m	-7 $^{\circ}$ 23 ^m	Oph
-8.0 ³	13.12	3.4	151	49	57	169	17.98	19.75	17 ^h 14 ^m	-3 $^{\circ}$ 34 ^m	Oph
-7.5 ²	12.65	3.3	181	31	75	198	9.72	22.22	16 ^h 49 ^m	-8 $^{\circ}$ 46 ^m	Oph
-6.3 ²	11.62	3.0	163	11	73	179	3.86	24.10	16 ^h 30 ^m	-12 $^{\circ}$ 03 ^m	Oph
-6.1 ³	11.48	2.9	121	19	47	131	8.92	20.99	16 ^h 52 ^m	-10 $^{\circ}$ 04 ^m	Oph
-5.0 ²	10.76	2.7	145	-5	69	161	-1.97	25.43	16 ^h 12 ^m	-15 $^{\circ}$ 19 ^m	Sco
-4.2 ³	10.21	2.6	97	-15	33	104	-8.79	18.58	16 ^h 16 ^m	-24 $^{\circ}$ 35 ^m	Sco
-3.8 ²	10.02	2.6	125	1	51	135	0.46	22.19	16 ^h 28 ^m	-15 $^{\circ}$ 40 ^m	Oph
-2.6 ²	9.37	2.4	95	-9	47	106	-5.41	26.22	16 ^h 01 ^m	-17 $^{\circ}$ 05 ^m	Lib
-2.3 ³	9.21	2.4	75	-49	17	91	-33.16	10.74	15 ^h 10 ^m	-45 $^{\circ}$ 35 ^m	Lup
-1.5 ²	8.81	2.3	83	-25	41	96	-16.76	25.31	15 ^h 32 ^m	-24 $^{\circ}$ 44 ^m	Lib

Stellar explosion times t_{SN} (before present) for the initial masses m of the supernova progenitors. The subgroups of the Sco-Cen association are identified by superscript numbers (UCL = 2, LCC = 3). Also listed are the mass-dependent ^{60}Fe yields, M_{ej} , (see also Extended Data Fig. 1) and the position coordinates (x, y, z) as defined by the local standard of rest. For each position a distance D is given, the corresponding galactic coordinates (longitude l and latitude b), the equatorial coordinates (right ascension α and declination δ in ICRS coordinates, J2000) and the stellar constellations (sc) in which the supernovae explosion sites would be observable at the present night sky.

Coherent feedback control of a single qubit in diamond

Masashi Hirose¹ & Paola Cappellaro¹

Engineering desired operations on qubits subjected to the deleterious effects of their environment is a critical task in quantum information processing, quantum simulation and sensing. The most common approach relies on open-loop quantum control techniques, including optimal-control algorithms based on analytical¹ or numerical² solutions, Lyapunov design³ and Hamiltonian engineering⁴. An alternative strategy, inspired by the success of classical control, is feedback control⁵. Because of the complications introduced by quantum measurement⁶, closed-loop control is less pervasive in the quantum setting and, with exceptions^{7,8}, its experimental implementations have been mainly limited to quantum optics experiments. Here we implement a feedback-control algorithm using a solid-state spin qubit system associated with the nitrogen vacancy centre in diamond, using coherent feedback⁹ to overcome the limitations of measurement-based feedback, and show that it can protect the qubit against intrinsic dephasing noise for milliseconds. In coherent feedback, the quantum system is connected to an auxiliary quantum controller (ancilla) that acquires information about the output state of the system (by an entangling operation) and performs an appropriate feedback action (by a conditional gate). In contrast to open-loop dynamical decoupling techniques¹⁰, feedback control can protect the qubit even against Markovian noise and for an arbitrary period of time (limited only by the coherence time of the ancilla), while allowing gate operations. It is thus more closely related to quantum error-correction schemes^{11–14}, although these require larger and increasing qubit overheads. Increasing the number of fresh ancillas enables protection beyond their coherence time. We further evaluate the robustness of the feedback protocol, which could be applied to quantum computation and sensing, by exploring a trade-off between information gain and decoherence protection, as measurement of the ancilla–qubit correlation after the feedback algorithm voids the protection, even if the rest of the dynamics is unchanged.

To demonstrate coherent feedback with spin qubits, we choose two of the most common tasks for qubits—implementing the no-operation (NOOP) and NOT gates, while cancelling the effects of noise. A simple, measurement-based feedback scheme, exploiting one ancillary qubit, was proposed in ref. 15. The correction protocol (Fig. 1a) works by entangling the qubit–ancilla system before the desired gate operation. By selecting an entangling operation U_c appropriate for the type of bath acting on the system, information about the noise action is encoded in the ancilla state. After undoing the entangling operation, the qubit coherence can be restored by a feedback action, that is, by measuring the ancilla and applying a correction operation U_q on the qubit based on the measurement result. Although this scheme applies to a broad class of environments, in the case of dephasing noise, the existence of the appropriate unitaries, $U_c = \sigma_x$ and $U_q = \sigma_z$ (where σ_α are Pauli matrices), has been proven¹⁵. The principle of deferred measurements¹⁶ allows postponing measurements until the end of a quantum algorithm. Here it enables replacing the ancilla measurement

and classical feedback by coherent conditional gates. The final qubit state is then completely decoupled from the bath effects, which are instead imposed on the ancilla. Because the scheme is compatible with any gate that commutes with U_c , it is more flexible than dynamical decoupling techniques. Another advantage is that the scheme is valid independently of the characteristic timescale of the bath and thus it can be used even in the presence of a Markovian bath for which dynamical decoupling fails. These two advantages enable the use of the feedback scheme for sensing, for example to detect transverse magnetic fields or, by a simple modification of the algorithm, to protect the sensing qubit from bit-flip errors while detecting fields aligned with the z quantization axis^{17–19}.

We demonstrated experimentally this feedback-based protection algorithm with a spin system associated with a single nitrogen-vacancy (NV) centre in diamond. This system has emerged as a good candidate for quantum information processing, because its electronic spin-1 can be optically polarized and measured, and presents long coherence times even at room temperature. Here we use an auxiliary qubit associated with the ^{14}N nuclear spin-1 of the nitrogen vacancy, to protect the coherence of the NV electronic spin against dephasing noise. The Hamiltonian of the electronic–nuclear spin system is

$$\mathcal{H} = \Delta S_z^2 + \omega_e S_z + Q I_z^2 + \omega_n I_z + A S_z I_z + B(S_x I_x + S_y I_y)$$

where S and I are the electron and nuclear spin operators, respectively, $\Delta = 2.87$ GHz is the electronic zero-field splitting and $Q = -4.95$ MHz is the nuclear quadrupolar interaction. The two spins are coupled by an isotropic hyperfine interaction with a longitudinal component $A = -2.16$ MHz and a transverse component²⁰ $B = -2.62$ MHz that can be neglected to first order. A magnetic field is applied along the NV crystal axis $[111]$ to lift the degeneracy of the $m_S = \pm 1$ level, yielding the electron and nuclear Zeeman frequencies ω_e and ω_n , respectively. In the experiment, we use a subspace of the total system representing a two-qubit system. For convenience, we choose the space spanned by the four levels $\{|m_S, m_I\rangle = |0, 1\rangle, |0, 0\rangle, |-1, 1\rangle, |-1, 0\rangle\}$. The effective Hamiltonian can then be rewritten by using spin-1/2 operators and simplified by transforming to a rotating frame at the resonant frequency of the qubit ($\Delta - \omega_e$) and the ancilla ($\omega_n + Q$): $\mathcal{H}' = A(-\sigma_z^e + \sigma_z^n - \sigma_z^e \sigma_z^n) / 4$.

In our high-purity diamond sample, decoherence processes of the electronic spin can be mainly attributed to a bath of spin-1/2 ^{13}C nuclei in the lattice (1.1% natural abundance), yielding a short dephasing time $T_{2e}^* \approx 4$ μs . We neglect the interaction between the bath and the ancillary spin, which couples very weakly to any source of magnetic field noise; indeed, the much longer coherence time of the ^{14}N spin, $T_{2n}^* \approx 3$ ms, is limited by the NV electronic spin relaxation and not by the ^{13}C bath.

In the experiment, the electronic spin qubit is initialized into the $m_S = 0$ state by optical excitation at 532 nm. At the chosen magnetic field strengths (390–510 G), optical pumping also polarizes the nuclear spin into the $m_I = +1$ state, owing to resonant polarization exchange

¹Research Laboratory of Electronics and Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

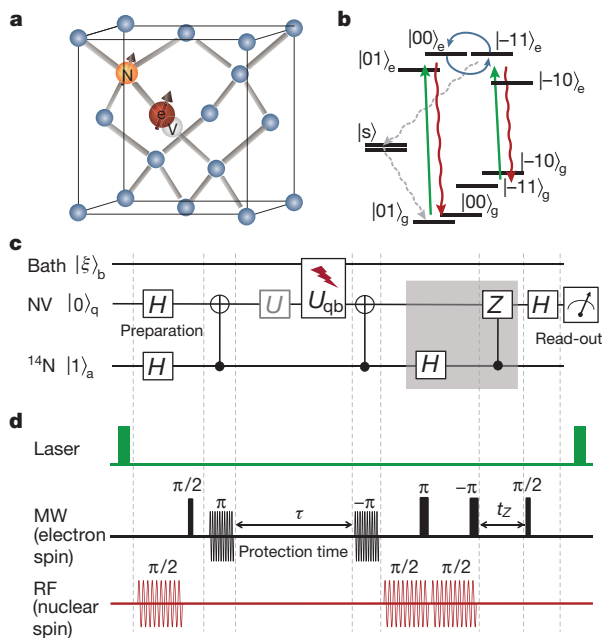


Figure 1 | Feedback algorithm and experimental implementation.

a, b, Nitrogen-vacancy (NV) centre in diamond (**a**) and relevant energy levels of the spin system (**b**), showing polarization processes under optical illumination²¹. **c**, Quantum circuit. Hadamard gates (H) prepare and read out a superposition state of the qubit, $|\phi\rangle_q = (|0\rangle + |1\rangle)/\sqrt{2}$. Amid entangling gates between qubit ('q') and ancilla ('a'), the qubit is subjected to noise (lightning bolt) and possibly unitary gates U . We assume the ancilla is not affected by the bath ('b'), yielding $\mathbb{I}_a \otimes U_{qb}(\tau)$, with $U_{qb}(\tau)$ the qubit–bath joint evolution. Given a dephasing bath, we set the entangling gate to $U_c = \sigma_x$ (conditional-NOT gate). More generally, upon undoing the entangling operation, the system is left in the state $|\Psi(\tau)\rangle = (|0\rangle_a K^+ |\phi_q, \xi_b\rangle + |1\rangle_a K^- |\phi_q, \xi_b\rangle)/\sqrt{2}$, with $K^\pm = U_{qb} \pm U_c U_{qb}^\dagger$. The entangling gate U_c is designed such that $K^+ = \mathbb{I}_q \otimes \chi_b^+$ and $K^- = U_q^\dagger \otimes \chi_b^-$, in which χ_b^\pm act on the bath only, and U_q acts on the qubit only. After measuring the ancilla, we could use a feedback operation U_q to restore the correct qubit state. The ancilla measurement is replaced by coherent feedback (shaded region) obtained by a controlled-correction gate (here $U_q = \sigma_z$ (Z) for dephasing noise). The final state of the combined system is then $(|0\rangle_a \chi_b^+ + |1\rangle_a \chi_b^-) |\phi_q, \xi_b\rangle/\sqrt{2}$, which reveals how the qubit is now decoupled from the bath. **d**, Experimental implementation. The laser excitation polarizes both spins. Black sinusoidal lines refer to selective microwave (MW) pulses acting only in the $m_I = 1$ manifold (thus mimicking controlled-NOT gates); solid bars indicate non-selective pulses. The radio-frequency (RF) excitation describes selective pulses in the $m_s = 0$ manifold. We use $\pi/2$ rotations about x to approximate Hadamard gates. To implement a non-selective radio-frequency $\pi/2$ gate on the nuclear spin we embed a non-selective microwave π pulse into two consecutive radio-frequency $\pi/2$ pulses. The controlled-correction gate is implemented by free evolution (t_z) under hyperfine coupling.

with the electronic spin in the excited state²¹. After initializing the qubit in the superposition state $|\phi\rangle_q = (|0\rangle_e + |1\rangle_e)/\sqrt{2}$, we implement the feedback-based protection algorithm using the control operations shown in Fig. 1b. Gate operations on the qubit and ancillary spins are performed by external microwave and radio-frequency fields²², respectively (see Methods). Because the noise is dephasing in nature, the feedback-based protection algorithm requires the entangling gate to be a controlled-NOT gate, $U_c = \sigma_x$, while the coherent feedback gate is implemented by a controlled Pauli-Z gate, $U_q = \sigma_z$ (see Fig. 1a and Methods). Finally, the qubit state is optically read out by monitoring spin-state dependent fluorescence.

We measure the fidelity of the NOOP and NOT gates while varying the protection time τ . As shown in Fig. 2a, the feedback algorithm protects the qubit coherence against the noise created by the ^{13}C nuclear spin bath for times $\tau > 1$ ms, much longer than the dephasing time. The spin bath in this experimental system is non-Markovian

and so dynamical decoupling techniques would reach similar protection times, as previously demonstrated²³; however, unlike dynamical decoupling, this method would also be effective in the Markovian noise regime, with its ultimate limit set only by the coherence time of the ancillary spin and the qubit relaxation. This limit could be overcome by re-initializing the ancilla or, if additional ancillas (such as nearby ^{13}C spins) are available, by repeating the algorithm with fresh ancillas, or by concatenating layers of protections against multi-axis noise (see Methods). Moreover, our scheme is more flexible than other protection schemes, including coherent transfer to the ancilla qubit²⁴, as it allows the application of some gate operations on the qubit¹⁵. For example, in our implementation, the qubit was evolving under the action of the hyperfine coupling, as indicated by the coherent oscillations of the fidelity (these oscillations could be removed by an adequate timing choice). In addition, we can implement a gate on the qubit, for example a NOT gate, during the feedback-based protection algorithm (Fig. 2d), without degrading the protection, even at times much longer than the gate itself and independent of the time at which the gate is performed (see Methods).

The success of the feedback-based protection algorithm relies on the fact that the increased qubit entropy, due to the coupling to the environment, is dumped on the ancilla. This entropy transfer could be revealed by measuring the state of the ancilla, which would yield information about the noise, while preserving the qubit state. If, instead, information about the ancilla is collected by a correlated qubit–ancilla measurement, the protection fails²⁵, as would happen in a measurement-based feedback for an imperfect ancilla read-out. Here we investigate this trade-off between the protection power and information gain on the ancilla in two different experiments. First, we transfer a part of the ancilla entropy back to the qubit by using a conditional gate that maps the state of the ancilla onto the qubit, correlating the two qubit states. By changing the angle of rotation of the conditional gate we vary the strength of the ancilla measurement from a weak to a strong one. As more information about the ancilla is acquired, the fidelity of the protection gate decreases (Fig. 2b). Note that we can combine the conditional gate performing the weak measurement of the ancilla state with the last conditional Z-gate of the algorithm, thus in practice performing a conditional phase-shift gate. The qubit fidelity is maximized when the conditional gate performs the required π rotation on the qubit, whereas it decreases when a different phase rotation is used.

Although this experiment clearly shows that the effectiveness of the protection is lost when a measurement of the ancilla–qubit correlation is performed, we can show that similar results also arise naturally from non-ideal experimental measurement conditions. We thus repeated the experiment at a different magnetic field, for which our signal—the fluorescence intensity—carries information about the state of both the qubit and the ancilla, and their correlation. At low magnetic field, we can approximate the normalized measurement operator by $\bar{M}_0 = |0\rangle\langle 0| \otimes \mathbb{I}$; however, for magnetic fields close to the excited-state level anti-crossing ($B \approx 500$ G such that the Zeeman energy ω_e is close to the excited-state zero-field splitting $\Delta_e \approx 1.4$ GHz) the normalized observable is $\bar{M}_0 - (\epsilon|10\rangle\langle 10| + \eta|00\rangle\langle 00|)$, where ϵ and η account for the fact that each electronic–nuclear spin level considered has a different fluorescence emission intensity due to the excited state dynamics²⁶.

Because the measurement observable at this magnetic field contains partial information about correlations between the qubit and ancilla states, the qubit fidelity at long times is reduced (Fig. 2a), reflecting the effects of the noise acting on the qubit. We thus reveal a trade-off between the amount of information acquired about the ancillary system and the protection that it provides to the qubit. In addition, we single out a measurement of the ancilla alone, without qubit–ancilla correlations, by comparing the signal acquired so far with the signal after the qubit is rotated by a π pulse just before detection. If the measurements were independent of the state of the ancilla, the average of

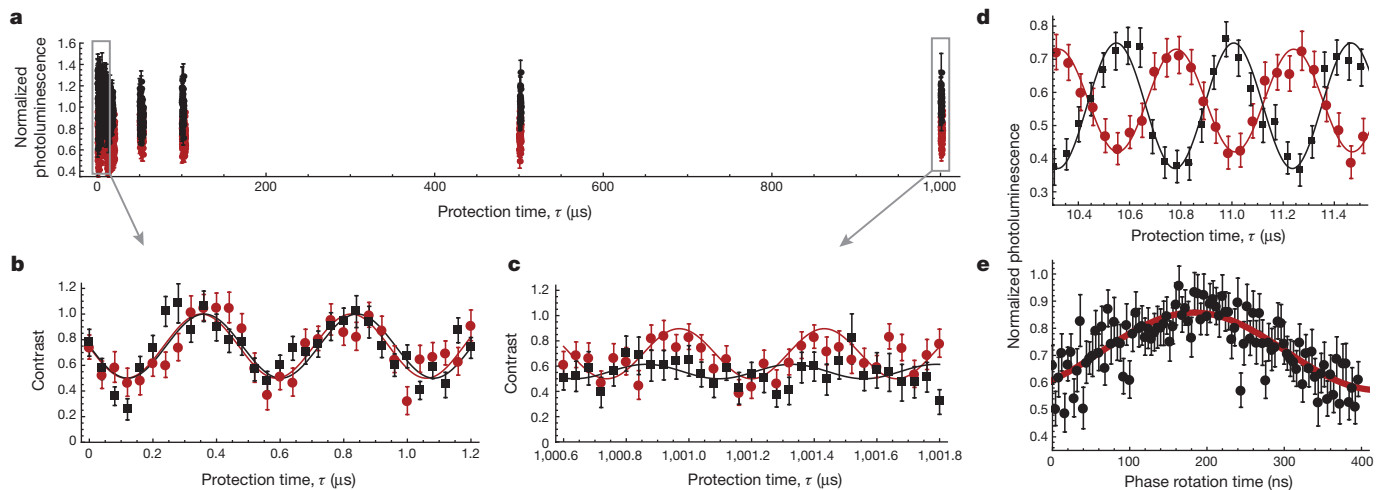


Figure 2 | Experimental demonstration of the feedback-based protection algorithm. **a**, The signal—normalized photoluminescence intensity—oscillates at the hyperfine coupling frequency, $A = -2.16$ MHz. The initial coherent superposition state of the qubit is preserved for a time $\tau > 1$ ms at $B = 390$ G (red circles), while we observe a sharp decrease in the signal amplitude at $B = 514$ G (black squares), where correlations between the qubit and ancilla states are partly measured. **b**, **c**, This behaviour is evident when we compare the fidelity at short (**b**) and long (**c**) times, for $B = 390$ G (red circles) and $B = 514$ G (black squares). To highlight the differences while taking into account different photoluminescence intensities at the two fields, we normalized all the data so that at short times the signal has the same (maximum) contrast. **d**, Protected NOT gate (red circles). The coherence of the qubit is

protected for a time longer than the dephasing time ($\tau > T_{2e}^*$) even when a NOT gate is applied. We compare the dynamics to the NOOP dynamics (black squares), demonstrating that the NOT gate inverts the state of the qubit, as indicated by the out-of-phase oscillations. **e**, Weak measurement of the ancilla. Normalized photoluminescence signal after a protection time $\tau = 8$ μ s, as a function of the strength of the ancilla measurement. In the experiment, we vary the strength of the ancilla measurement by changing the angle of the last controlled phase rotation gate by varying its time. All error bars represent the standard deviation of the signal, calculated by error propagation from the photoluminescence intensity of the signal and of reference photoluminescence curves acquired for each data point for $m_S = \{0, -1\}$. The lines are fits to the model presented in Methods.

the two experiments should not vary in time (because it corresponds to measuring the identity operator); indeed, this is what we observed in experiments performed at the lower magnetic field (Fig. 3a, c). By contrast, for a magnetic field close to the level anti-crossing, the presence of oscillations indicates that we can extract information about the ancilla. However, at times longer than the qubit dephasing time, the state of the ancilla has completely decohered (Fig. 3b, d), thus quenching these oscillations.

In conclusion, we performed coherent feedback control on a single solid-state qubit associated with the NV centre in diamond. The feedback algorithm was applied to protect the qubit coherence, while performing two essential qubit gates—NOOP and NOT gates—during

the protection time. We showed that this feedback-based protection algorithm protects the qubit coherence far beyond the dephasing time of the qubit, even if no active control is applied to decouple it from the noise. The algorithm can be extended to applications in quantum information processing and quantum sensing, and it could be implemented in many other hybrid spin systems, such as phosphorus²⁷ or antimony²⁸ donors in silicon, defects in silicon carbide²⁹, or quantum dots³⁰. Because we applied a coherent feedback protocol, thus avoiding measurement of the state of the ancilla, the decoherent effects of the bath are effectively stored in the ancilla. Consequently, we were able to explore an interesting compromise between the amount of information gained about the ancilla and the effectiveness of the protection

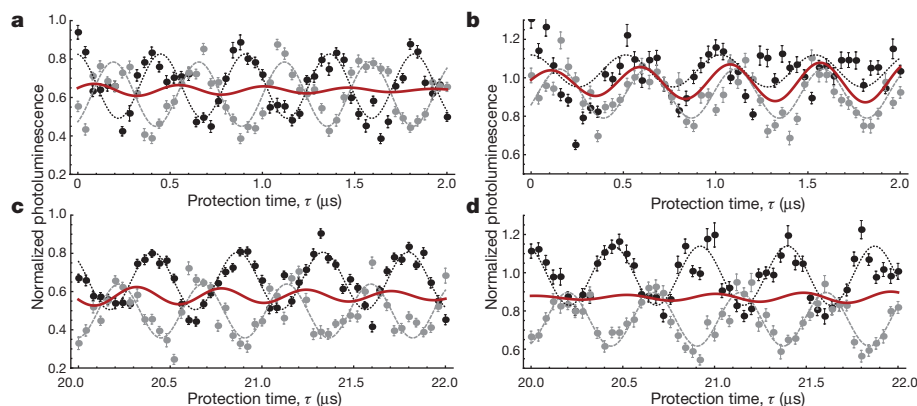


Figure 3 | Partial measurement of the ancilla. **a–d**, Comparison of the fidelity signal with (grey) and without (black) a π pulse on the qubit, revealing the amount of information acquired about the ancilla qubit state. Signals were measured at short protection times, $\tau < T_{2e}^*$ (0–2 μ s; **a**, **b**), and at longer times, $\tau > T_{2e}^*$ (20–22 μ s; **c**, **d**). Error bars represent the standard deviation of the signal (see Fig. 2). The data are fitted (lines) using the model described in Methods. At the lower magnetic field ($B = 390$ G; **a**, **c**), the average signals (red lines, obtained from the fits to

the data) exhibit only weak oscillations, indicating that at this field the measurement carries very little information about the ancilla state. For magnetic fields close to the level anti-crossing ($B = 514$ G; **b**, **d**), the oscillation of the average signal is more pronounced and is observed until $\tau < T_{2e}^*$, whereas it disappears at longer times. This signal behavior is an indication that the ancillary spin effectively decoheres on the T_{2e}^* timescale as a result of the feedback algorithm.

algorithm. These results contribute to elucidating the robustness of feedback control and to paving the way for a more widespread adoption of close-loop control strategies for solid-state qubit systems.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August 2015; accepted 3 February 2016.

1. Boscain, U. & Mason, P. Time minimal trajectories for a spin 1/2 particle in a magnetic field. *J. Math. Phys.* **47**, 062101 (2006).
2. Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbrüggen, T. & Glaser, S. J. Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *J. Magn. Reson.* **172**, 296–305 (2005).
3. Wang, X. & Schirmer, S. G. Analysis of Lyapunov method for control of quantum states. *Automatic Control. IEEE Trans. Automatic Control* **55**, 2259–2270 (2010).
4. Ajoy, A. & Cappellaro, P. Quantum simulation via filtered Hamiltonian engineering: application to perfect quantum transport in spin networks. *Phys. Rev. Lett.* **110**, 220503 (2013).
5. Wiseman, H. M. & Milburn, G. J. Quantum theory of optical feedback via homodyne detection. *Phys. Rev. Lett.* **70**, 548–551 (1993).
6. Doherty, A. C., Jacobs, K. & Jungman, G. Information, disturbance, and Hamiltonian quantum feedback control. *Phys. Rev. A* **63**, 062306 (2001).
7. Vijay, R. *et al.* Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback. *Nature* **490**, 77–80 (2012).
8. Sayrin, C. *et al.* Real-time quantum feedback prepares and stabilizes photon number states. *Nature* **477**, 73–77 (2011).
9. Lloyd, S. Coherent quantum feedback. *Phys. Rev. A* **62**, 022108 (2000).
10. Viola, L., Knill, E. & Lloyd, S. Dynamical decoupling of open quantum systems. *Phys. Rev. Lett.* **82**, 2417–2421 (1999).
11. Shor, P. W. Fault-tolerant quantum computation. In *37th Annual Symposium on Foundations of Computer Science* 56–65 (IEEE Comput. Soc., 1996).
12. Taminiau, H. T., Cramer, J., van der Sar, T., Dobrovitski, V. V. & Hanson, R. Universal control and error correction in multi-qubit spin registers in diamond. *Nature Nanotechnol.* **9**, 171–176 (2014).
13. Waldherr, G. *et al.* Quantum error correction in a solid-state hybrid spin register. *Nature* **506**, 204–207 (2014).
14. Cramer, J. *et al.* Repeated quantum error correction on a continuously encoded qubit by real-time feedback. Preprint at <http://arXiv.org/abs/1508.01388> (2015).
15. Ticozzi, F. & Viola, L. Single-bit feedback and quantum-dynamical decoupling. *Phys. Rev. A* **74**, 052328 (2006).
16. Griffiths, R. & Niu, C.-S. Semiclassical Fourier transform for quantum computation. *Phys. Rev. Lett.* **76**, 3228–3231 (1996).
17. Kessler, E. M., Lovchinsky, I., Sushkov, A. O. & Lukin, M. D. Quantum error correction for metrology. *Phys. Rev. Lett.* **112**, 150802 (2014).
18. Arrad, G., Vinkler, Y., Aharonov, D. & Retzker, A. Increasing sensing resolution with error correction. *Phys. Rev. Lett.* **112**, 150801 (2014).
19. Unden, T. *et al.* Quantum metrology enhanced by repetitive quantum error correction. Preprint at <http://arXiv.org/abs/1602.07144> (2015).
20. Chen, M., Hirose, M. & Cappellaro, P. Measurement of transverse hyperfine interaction by forbidden transitions. *Phys. Rev. B* **92**, 020101 (2015).
21. Jacques, V. *et al.* Dynamic polarization of single nuclear spins by optical pumping of nitrogen-vacancy color centers in diamond at room temperature. *Phys. Rev. Lett.* **102**, 057403 (2009).
22. Laraoui, A. *et al.* High-resolution correlation spectroscopy of ^{13}C spins near a nitrogen-vacancy centre in diamond. *Nature Commun.* **4**, 1651 (2013).
23. Ryan, C. A., Hodges, J. S. & Cory, D. G. Robust decoupling techniques to extend quantum coherence in diamond. *Phys. Rev. Lett.* **105**, 200402 (2010).
24. Morton, J. J. L. *et al.* Solid-state quantum memory using the ^{31}P nuclear spin. *Nature* **455**, 1085–1088 (2008).
25. Kim, Y.-H., Yu, R., Kulik, S. P., Shih, Y. & Scully, M. O. Delayed “choice” quantum eraser. *Phys. Rev. Lett.* **84**, 1–5 (2000).
26. Steiner, M., Neumann, P., Beck, J., Jelezko, F. & Wrachtrup, J. Universal enhancement of the optical readout fidelity of single electron spins at nitrogen-vacancy centers in diamond. *Phys. Rev. B* **81**, 035205 (2010).
27. Pla, J. J. *et al.* High-fidelity readout and control of a nuclear spin qubit in silicon. *Nature* **496**, 334–338 (2013).
28. Wolfowicz, G. *et al.* Conditional control of donor nuclear spins in silicon using stark shifts. *Phys. Rev. Lett.* **113**, 157601 (2014).
29. Widmann, M. *et al.* Coherent control of single spins in silicon carbide at room temperature. *Nature Mater.* **14**, 164–168 (2015).
30. Chekhovich, E. A. *et al.* Nuclear spin effects in semiconductor quantum dots. *Nature Mater.* **12**, 494–504 (2013).

Acknowledgements We thank L. Viola, F. Ticozzi, M. Lukin and F. Jelezko for discussions. This work was supported in part by the US Air Force Office of Scientific Research grant number FA9550-12-1-0292 and by the US Office of Naval Research grant number N00014-14-1-0804.

Author Contributions M.H. and P.C. contributed to all aspects of this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (pcappell@mit.edu).

METHODS

Measurement-based versus coherent feedback. In feedback protocols, the goal is to control an open system (the plant) so that it undergoes a desired evolution or it reaches a desired state. To achieve this goal, feedback protocols engineer a second system (the controller) connected to the plant. In quantum feedback there are two possible strategies. In the first strategy, we measure the plant output and engineer a classical controller that manipulates the plant according to the measurement result (this is called a measurement-based feedback protocol). In the second strategy, the controller is a second quantum system, coherently coupled to the plant. This strategy, called coherent feedback⁹, allows avoiding measurements. The controller can gain information about the plant state by coherent operations, sharing information about the state via entangling operations. Then, the controller can manipulate the plant on the basis of this information to achieve the desired plant dynamics, by performing conditional operations on the plant.

Although the first strategy is more intuitive and closer to the classical scenario, coherent feedback is more intriguing, both from a theoretical point of view, owing to its relation to information and entanglement theory, and from an experimental one, because avoiding measurements often simplifies the experimental implementation. In addition, coherent feedback formally encompasses measurement-based feedback and, as we did here, a measurement-based protocol can be transformed into a coherent one. To clarify the two strategies, we present the feedback protocol we considered¹⁵ using both strategies.

In both protocols, the ancillary qubit is first prepared in a superposition state, so that the initial state of the combined system is $|\Psi_0\rangle = (|0\rangle_a + |1\rangle_a)|\phi\rangle_q|\xi\rangle_b/\sqrt{2}$, where $|\xi\rangle_b$ and $|\phi\rangle_q$ are the initial state of the bath and qubit, respectively. The qubit and the ancilla are then entangled by a controlled operation U_c , $|\Psi_1\rangle = (|0\rangle_a|\phi\rangle_q + |1\rangle_a U_c|\phi\rangle_q)|\xi\rangle_b/\sqrt{2}$. The choice of U_c depends on the type of noise to be refocused. As explained in ref. 15, U_c can be chosen so that it transforms the qubit–bath interaction into a block-diagonal form. This enables performing appropriate correction operations in each sub-manifold of the qubit. More intuitively, if the action of the bath can be represented by a random unitary operator $U_{qb} = e^{i\phi_r \sigma_n}$, with ϕ_r random and \mathbf{n} a unit vector, then U_c should invert the action of the bath: $U_c^\dagger U_{qb} U_c = U_{qb}^\dagger$. After the system evolves under the action of the bath (and possibly of a desired gate), the entangling operation is undone, $|\Psi_2\rangle = (|0\rangle_a U_{qb}|\phi\rangle_q|\xi\rangle_b + |1\rangle_a U_c^\dagger U_{qb} U_c|\phi\rangle_q|\xi\rangle_b)/\sqrt{2}$. Owing to the entanglement, information about the action of the bath is encoded in the state of the ancilla. Thus, in the measurement-based feedback algorithm, the ancilla state is measured (in the x basis) and, depending on the outcome, a classical control operation is applied to the qubit. We can rewrite the state just before the measurement as

$$|\Psi_2\rangle = \frac{1}{2} \left[|+\rangle (U_{qb} + U_c^\dagger U_{qb} U_c) + |-\rangle (U_{qb} - U_c^\dagger U_{qb} U_c) \right] |\phi\rangle_q |\xi\rangle_b$$

where $|\pm\rangle$ are eigenstates of σ_x^a . With an appropriate choice of U_c , $K^\pm = U_{qb} \pm U_c^\dagger U_{qb} U_c$ reduce to $K^+ = \mathbb{1}_q \otimes \chi_b^+$ and $K^- = U_q^\dagger \otimes \chi_b^-$, where χ_b^\pm are operators acting on the bath only and U_q is a unitary acting on the qubit. If the measurement of the ancilla yields the negative eigenstate, then we apply the correction operation U_q , which is a unitary that depends on the type of bath that affects the qubit. In the simple noise model we considered above (unitary noise with a stochastic Hamiltonian generator $H_r \propto \sigma \cdot \mathbf{n}$), the correction operation is $U_q = i\sigma \cdot \mathbf{n}$.

In the experimental implementation, the main source of decoherence is dephasing noise; therefore, we fixed $U_c = \sigma_x$ and $U_q = \sigma_z$. In a magnetometry scenario, when we want to correct for bit-flip errors while measuring longitudinal external fields, the gates would be $U_c = \sigma_z$ and $U_q = \sigma_x$.

In the coherent feedback strategy, the measurement and classical correction operation are replaced by a coherent conditional operation (see Extended Data Fig. 1a, b). The measurement (that is, conveying information to an external observer) is not needed for the ancilla to be able to perform a correction operation, because the ancilla already encodes the necessary information about the qubit state and the action of the noise.

At the end of the feedback protocol, the ancilla can be recycled to perform a second round of protection, thus allowing the protection to be extended beyond the coherence time of the ancilla. In the current experiment, we could achieve the first strategy by using coherent polarization exchange between the nitrogen vacancy (NV) and ^{14}N spins at lower (or much higher) magnetic field, at which optical polarization leaves the nuclear spin untouched. After a first run of the algorithm (for a protection time $\tau < T_{2n}^*$), we could swap the state of the NV electronic and nuclear spins, re-polarize the NV centre and swap back the state from the nuclear spin to the NV, before applying the feedback protection for another stretch of time. Alternatively, if the ancilla cannot be re-initialized, then we could use additional fresh ancillas (see Extended Data Fig. 1c), for example, other ^{13}C nuclear spins in the lattice. If two ancillas are available, then we can construct a concatenated feedback algorithm (see Extended Data Fig. 1d) that

protects the qubit against multi-axis noise. The qubit is then protected against, for example, both dephasing (T_2) and depolarizing (T_1) noise. More broadly, the feedback-based protection algorithm could provide a first layer of protection against the strongest noise source, and be combined with more advanced quantum error-correcting codes (even on the ancilla itself) with the goal of achieving fault tolerance.

Modelling of the time evolution of the qubit–ancilla system. We study the time evolution of the qubit during the algorithm under the effects of a bath, to better understand the action of the feedback-based protection algorithm and to obtain insight into its limits. Although in the experiment the spin bath is non-Markovian, we can use a simpler model with a Markovian bath to find an analytical solution, because it yields the same result up to a different form of the exponential decays.

In the model, we consider dephasing of the qubit and the ancilla, and the lattice relaxation of the qubit, characterized by the timescales T_{2e}^* , T_{2n}^* and T_{1e} , respectively. Letting ρ be a state of the qubit and the ancilla, the time evolution under the Markovian environment is determined by the Lindblad equation³¹

$$\frac{d\rho}{dt} = i[\rho, \mathcal{H}] + \sum_k \left(L_k \rho L_k^\dagger - \frac{1}{2} \rho L_k^\dagger L_k - \frac{1}{2} L_k^\dagger L_k \rho \right)$$

where

$$L_1 = \sqrt{\frac{1}{2T_{2e}^*}} \sigma_z^e, \quad L_2 = \sqrt{\frac{1}{T_{1e}}} \sigma_z^e, \quad L_3 = \sqrt{\frac{1}{T_{1e}}} \sigma_z^e, \quad L_4 = \sqrt{\frac{1}{2T_{2n}^*}} \sigma_n^e$$

Given an initial state $\rho_0 = |0\rangle\langle 0|_e \otimes |1\rangle\langle 1|_n$, the state after the NOOP gate for a protection time τ is

$$\begin{aligned} \rho(\tau) = & \frac{1}{4} \mathbb{1} - \frac{1}{4} \exp\left(-\frac{\tau}{T_{1e}}\right) \cos^2\left(\frac{A\tau}{2}\right) \left\{ \exp\left[-\tau\left(\frac{1}{4T_{2e}^*} + \frac{1}{4T_{2n}^*}\right)\right] \sigma_z^n \right. \\ & + \exp\left(-\frac{\tau}{4T_{2n}^*}\right) \sigma_z^e + \exp\left(-\frac{\tau}{4T_{2e}^*}\right) \sigma_z^e \sigma_n^n \Big\} \\ & + \frac{1}{8} \exp\left[-\tau\left(\frac{1}{T_{1e}} + \frac{1}{4T_{2e}^*} + \frac{1}{4T_{2n}^*}\right)\right] \sin(A\tau) (\sigma_x^n - \sigma_x^e \sigma_x^n) \\ & + \frac{1}{8} \exp\left[-\tau\left(\frac{1}{T_{1e}} + \frac{1}{4T_{2n}^*}\right)\right] \sin(A\tau) (-\sigma_y^e + \sigma_y^e \sigma_y^n) \\ & + \frac{1}{4} \exp\left(-\frac{\tau}{T_{1e}}\right) \sin^2\left(\frac{A\tau}{2}\right) \left\{ \exp\left[-\tau\left(\frac{1}{4T_{2e}^*} + \frac{1}{4T_{2n}^*}\right)\right] \sigma_x^e \sigma_z^n \right. \\ & + \exp\left(-\frac{\tau}{4T_{2n}^*}\right) \sigma_z^e \sigma_y^n + \exp\left(-\frac{\tau}{4T_{2e}^*}\right) \sigma_y^e \sigma_x^n \Big\} \end{aligned}$$

The effectiveness of the feedback-based protection algorithm can be evaluated by measuring the probability of retaining the initial qubit state. The state fidelity is then given by the ideal normalized measurement operator $M_0 = |0\rangle\langle 0|_e$ that gives the measured signal \bar{S}

$$\bar{S}(\tau) = \text{Tr}[\rho(\tau) M_0] = \frac{1}{2} - \frac{1}{2} \cos^2\left(\frac{A\tau}{2}\right) \exp\left[-\tau\left(\frac{1}{T_{1e}} + \frac{1}{4T_{2n}^*}\right)\right]$$

This result clearly indicates that the fidelity is not limited by the dephasing of the qubit T_{2e}^* , but by that of the ancillary spin and the T_{1e} relaxation time of the qubit.

If instead we measure the normalized observable at the magnetic field corresponding to the excited-state level anti-crossing (ESLAC), then we gather some information about the ancilla, resulting in a degradation of the protection

$$\begin{aligned} \bar{S}(\tau) = & \text{Tr}[\rho(\tau) \bar{M}] \\ = & \frac{1}{2} - \frac{1}{4} (\epsilon + \eta) - \frac{1}{2} \cos^2\left(\frac{A\tau}{2}\right) \exp\left[-\tau\left(\frac{1}{T_{1e}} + \frac{1}{4T_{2n}^*}\right)\right] \\ & + \frac{1}{4} \cos^2\left(\frac{A\tau}{2}\right) \exp\left(-\frac{\tau}{T_{1e}}\right) \left\{ (\epsilon - \eta) \left[\exp\left(-\frac{\tau}{4T_{2e}^*}\right) + \exp\left(-\frac{\tau}{4T_{2n}^*}\right) \right] \right. \\ & \left. - (\epsilon + \eta) \exp\left[-\tau\left(\frac{1}{4T_{2e}^*} + \frac{1}{4T_{2n}^*}\right)\right] \right\} \end{aligned}$$

Although the first line still indicates that the state is protected by the presence of the ancilla, terms in the other lines, which decay at the rate of T_{2e}^* , originate from the measurement of the ancilla spin and can be observed as a reduction of the oscillations of signal in the experiment.

In the main text we showed that it is possible to perform a NOT gate during the protection time. Because the gate time is typically short (58 ns in our experiments), we studied the subsequent evolution of the qubit, to show that its coherence is preserved for times longer than the dephasing time, $T_{2e}^* \approx 4 \mu\text{s}$. In addition, it is possible to insert the NOT gate at any point during the protected evolution. Owing to the coupling with the ancilla qubit, the resulting signal is

$$\bar{S}_\pi(\tau, \tau_\pi) = \text{Tr}[\rho_\pi(\tau, \tau_\pi) \bar{M}_0] \\ = \frac{1}{2} + \frac{1}{4} [\cos(A\tau - A\tau_\pi) + \cos(A\tau_\pi)] \exp\left[-\tau\left(\frac{1}{T_{1e}} + \frac{1}{4T_{2n}^*}\right)\right] \quad (1)$$

where τ_π is the time at which the NOT gate is applied (with τ the total protection time). This expression was used to fit the experimental results obtained in the main paper and the data presented in Extended Data Fig. 2.

Weak measurement of the ancilla-qubit correlation. To observe the compromise between protection and information gained on the ancillary qubit, we perform a weak measurement of the nuclear-electronic spin correlation by applying a conditional gate after the feedback algorithm and before the detection of the NV electronic spin by optical read-out. As shown in Extended Data Fig. 3a, the weak measurement could be obtained by applying a conditional rotation of the NV electronic spin, after that spin had been rotated back to its population state by a $\pi/2$ pulse. The rotation angle ϕ determines the strength of the measurement.

However, these additional gates can be avoided by noting that the circuit is equivalent to replacing the conditional π Z-gate with a conditional θ Z-gate, where $\theta = \pi/2 - \phi$ sets the measurement strength; see Extended Data Fig. 3b.

The state after the feedback algorithm is

$$\rho_\theta(\tau) = \frac{1}{4} 1 - \frac{1}{4} \exp\left(-\frac{\tau}{T_{1e}}\right) \cos^2\left(\frac{A\tau}{2}\right) \left\{ \exp\left[-\tau\left(\frac{1}{4T_{2e}^*} + \frac{1}{4T_{2n}^*}\right)\right] \sigma_z^n \right. \\ \left. + \exp\left(-\frac{\tau}{4T_{2n}^*}\right) \sigma_z^e + \exp\left(-\frac{\tau}{4T_{2e}^*}\right) \sigma_z^e \sigma_z^n \right\} + \frac{1}{8} \alpha(\theta) \exp\left(-\frac{\tau}{T_{1e}}\right) \\ \times \cos^2\left(\frac{A\tau}{2}\right) \exp\left(-\frac{\tau}{T_{2n}^*}\right) - \exp\left(-\frac{\tau}{T_{2e}^*}\right) (\sigma_z^e + \sigma_z^e \sigma_z^n) + \dots$$

where $\alpha(\theta) = 1 + \cos(A\tau/2 + \theta)/\cos(A\tau/2)$. Assuming no contrast of the measurement of the nuclear spin ($\epsilon = \eta = 0$), the signal is

$$\bar{S}_\theta(\tau) = \frac{1}{2} - \frac{1}{2} \cos^2\left(\frac{A\tau}{2}\right) \exp\left[-\tau\left(\frac{1}{4T_{2e}^*} + \frac{1}{4T_{2n}^*}\right)\right] \\ + \frac{1}{2} \alpha(\theta) \cos^2\left(\frac{A\tau}{2}\right) \exp\left(-\frac{\tau}{T_{1e}}\right) \left[\exp\left(-\frac{\tau}{T_{2n}^*}\right) - \exp\left(-\frac{\tau}{T_{2e}^*}\right) \right]$$

Although for $\theta = \pi$ ($\alpha = 0$) we recover the ideal fidelity, protected against the electron dephasing noise, when $\theta \neq \pi$ we perform a partial measurement of the ancilla, but the signal now decays according to the time constant T_{2e}^* .

Experimental set-up. Experiments were performed using an electronic-grade single-crystal diamond plate provided by Element 6, with [100] orientation and a nominal nitrogen concentration of less than 5 parts per billion. A home-built confocal microscope was used to scan the sample (scanning piezo stage Nano-3D200, Mad City Labs) and to measure a single naturally occurring NV centre situated about $10 \mu\text{m}$ from the surface of the diamond. The NV centre is coherently excited by an optical pulse at 532 nm provided by a diode-pump laser (Coherent Compass 315M) and shaped by an acoustic optical modulator (AOM) with rise time ≤ 7 ns (1250C-848, Isomet), and then delivered by means of a $100\times$ oil immersion objective lens with numerical aperture $\text{NA} = 1.3$ (Nikon Plan Fluor). The fluorescence by the NV centre is collected by the same objective lens and then sent to a single-mode broadband fibre of $\text{NA} = 0.12$ (Font Canada), being connected to a single photon detector (SPCM-AQRH-13-FC, Perkin Elmer). An arbitrary waveform generator (AWG) with 1.2 GS s^{-1} (AWG5014B, Tektronix) plays a central role in generating microwave and radio-frequency pulses, as well as timing the AOM and the single-photon detector. Microwave pulses are shaped by mixing the signal from the AWG with a carrier microwave field provided by a signal generator (N5183A-520, Agilent) with the aid of an I/Q mixer (IQ-0318L, Marki Microwave), and then cleaned up by a switch (ZASWA-2-50DR+, Mini-Circuits); radio-frequency pulses

are directly generated by the AWG. The microwave pulse and radio-frequency pulses are combined after amplification (GT-1000A, Gigatronics for microwave and LZY-22+, Mini-Circuits for radio-frequency) and delivered to the sample via a thin copper wire with a diameter of $25 \mu\text{s}$ (Alfa Aesar).

Dynamics and control of the two-qubit system. *Hamiltonian of the NV- ^{14}N system.* The NV-centre spin system consists of the electronic spin ($S = 1$) and the nitrogen nuclear spin ($I = 1$) that interact strongly via the hyperfine coupling ($A \approx -2.16 \text{ MHz}$). A static magnetic field is applied to lift the degeneracy of the electronic $m_S = \pm 1$ and nuclear spin $m_I = \pm 1$ levels, thus yielding the Hamiltonian

$$\mathcal{H} = \Delta S_z^2 + \omega_e S_z + Q I_z^2 + \omega_n I_z + A S_z I_z$$

where S and I are the electron and nuclear spin operators, respectively, $\Delta = 2.87 \text{ GHz}$ is the electronic zero-field splitting and $Q = -4.95 \text{ MHz}$ is the nuclear quadrupolar interaction. Here $\omega_{e/n} = \gamma_{e/n} B$ are the Zeeman energies, with the gyromagnetic ratios $\gamma_e = 2.8 \text{ MHz G}^{-1}$ and $\gamma_n = -0.308 \text{ kHz G}^{-1}$ for the electronic and nuclear spin, respectively.

We define one qubit in the electronic spin subspace as $|m_S = 0\rangle \equiv |1\rangle_q$ and $|m_S = -1\rangle \equiv |0\rangle_q$. A second, ancillary qubit is defined in the subspace of the nuclear spin as $|m_I = +1\rangle \equiv |1\rangle_a$ and $|m_I = 0\rangle \equiv |0\rangle_a$. Both spins are driven on-resonance on these selected transitions; therefore, we can neglect other parts of the Hilbert space and express the Hamiltonian in the rotating frame in the subspace of interest. We then rewrite the Hamiltonian in terms of spin-1/2 operators $\sigma_{x,y,z}^{e,n}$ defined by the basis above as

$$\mathcal{H} = \frac{A}{4} (-\sigma_z^e + \sigma_z^n - \sigma_z^e \sigma_z^n)$$

The hyperfine interaction induces a frequency shift, thus enabling the selectivity of the microwave pulse to be adjusted by tuning the transition rate (Ω_e). A strong driving $\Omega_e = 20 \text{ MHz} \gg A$, is used to drive the qubit non-selectively for any ancilla state, whereas $\Omega_e = 500 \text{ kHz} \ll A$ is used to drive the qubit selectively on the ancilla state, thus engineering conditional gates. The radio frequency is on resonance with the transition between $|0, 1\rangle$ and $|0, 0\rangle$. The transition rate is about $\Omega_n = 20\text{--}40 \text{ kHz}$, which reflects an enhancement due to electronic virtual transitions mediated by the transverse component of the hyperfine coupling²⁰. Because the nuclear Rabi transition rate is always smaller than the detuning given by the hyperfine interaction ($\Omega_n \ll A$), only selective driving of the ancillary spin is available and unconditional gates need to be engineered with composite pulses.

Coherent control of the nuclear spin. The ^{14}N nuclear spin is coherently controlled by using a resonant radio-frequency field. In general, the nuclear spin state cannot be directly measured optically with high fidelity. Therefore, after driving the nuclear spin, we apply a gate to map the nuclear spin state to the NV electronic spin (Extended Data Fig. 4a). To calibrate the quantum gates required for the feedback-based protection algorithm ($\pi/2$ rotation), we first characterize the nuclear spin by measuring the resonance frequency of the transition $|1\rangle_q |1\rangle_a \leftrightarrow |1\rangle_q |0\rangle_a$ (Extended Data Fig. 4b) and the nuclear Rabi oscillations (Extended Data Fig. 4c). Despite the small gyromagnetic ratio of the nitrogen ($\gamma_n = 0.308 \text{ kHz G}^{-1}$), its Rabi oscillations are substantially enhanced, owing to the transverse hyperfine coupling with the NV spin (we achieve an enhancement factor of about 20 over the bare Rabi frequency around 500 G; ref. 20). To further characterize the nuclear spin ancillary qubit, we measured the dephasing time T_{2n}^* by performing a Ramsey experiment (Extended Data Fig. 4d) and obtained $T_{2n}^* \approx 3.2 \text{ ms}$.

Although the dynamics of the ^{14}N spin under radio-frequency driving is relatively simple when the NV electronic spin is in one of the population states, further complications arise when the NV qubit is in a superposition state. These complications stem from the electronic spin decoherence and its evolution due to the radio-frequency driving and the hyperfine interaction.

For a typical nuclear spin gate, such as the $\pi/2$ rotation required for the algorithm, we need to consider the decoherence process of the electronic spin, which progresses at a faster rate ($T_{2e}^* \approx 4 \mu\text{s}$) than the gate operation ($t_g \approx 30 \mu\text{s}$). In addition, we need to engineer unconditional gates, even when the driving field can only rotate the nuclear spin in one of the electronic spin manifolds. To protect the qubit state during the gate operation, a microwave π pulse is inserted in the middle of the radio-frequency pulse and is then compensated by adding a second microwave $-\pi$ pulse at the end of the radio-frequency pulse. This procedure simultaneously solves the second issue, because it produces a non-selective gate. The Hamiltonian under the radio-frequency driving in the rotating frame is $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{\text{RF}}$ with $\mathcal{H}_{\text{RF}}(t) = -\Omega_{e,z} \sigma_z^e \cos(\omega_{\text{RF}} t + \phi)/2 + \sqrt{2} \Omega_n \sigma_x^n$. Here $\Omega_{e,z}$ and Ω_n are the coupling strength of the electronic spin and the nuclear spin with the longitudinal and transverse component of the radio-frequency field, respectively. We can neglect the coupling of the nuclear spin with the longitudinal component of the radio-frequency field and the transverse component of the radio-frequency effects on the electronic field, because at most it induces a resonance shift (Bloch-Siegert

shift³²) that is refocused by the microwave π pulses embedded in the radio-frequency pulse. The longitudinal driving imposes an additional, quite large, phase modulation $\Theta(\tau)$ on the NV spin, even in the presence of the spin echo π pulses

$$\begin{aligned}\Theta(\tau) &= -\frac{\Omega_{e,z}}{2} \left[\int_0^\tau \sin(\omega_{\text{RF}}t + \phi) dt - \int_\tau^{2\tau} \sin(\omega_{\text{RF}}t + \phi) dt \right] \\ &= \frac{2\Omega_{e,z}}{\omega_{\text{RF}}} \cos(\omega_{\text{RF}}\tau + \phi) \sin\left(\frac{\omega_{\text{RF}}\tau}{2}\right)^2\end{aligned}\quad (2)$$

Although, in principle, we can remove this phase by a proper selection of the timing, in practice this often clashes with other time requirements set by the need to refocus effects due to the hyperfine couplings (explained below).

In our experiment, we used a more flexible strategy to remove this undesired modulation, which used time-proportional phase incrementation³³ (TPPI) of the radio-frequency phase $\phi(\tau)$. $\Theta(\tau)$ in equation (2) can be set to zero at any time by selecting $\phi(\tau) = -\omega_{\text{RF}}\tau$. Unlike stroboscopic detection with a time interval $\delta\tau = 2\pi/\omega_{\text{RF}}$, TPPI enables continuous measurement of the ^{14}N spin Rabi driving; it is also of great use when the time resolution of experiment system is limited and $\delta\tau$ cannot be selected with enough precision. Defining $U(\tau) = \exp(-i\mathcal{H}\tau)$ and $R_{\pm\pi} = \exp(\mp i\pi\sigma_x^e)$, the two-spin propagator $V(\tau)$ under TPPI conditions is

$$\begin{aligned}V(\tau) &= R_{-\pi}U(\tau)R_{\pi}U(\tau) \\ &= \exp\left\{-i\tau\left[\frac{A}{4}(\sigma_z^e + \sigma_z^n + \sigma_z^e\sigma_z^n) + \sqrt{2}\Omega_n\sigma_x^n\right]\right\} \\ &\quad \times \exp\left\{-i\tau\left[\frac{A}{4}(-\sigma_z^e + \sigma_z^n - \sigma_z^e\sigma_z^n) + \sqrt{2}\Omega_n\sigma_x^n\right]\right\} \\ &= \cos(\sqrt{2}\Omega_n\tau)\left[\cos\left(\frac{A\tau}{2}\right)\mathbb{1} - i\sin\left(\frac{A\tau}{2}\right)\sigma_z^n\right] \\ &\quad - i\sin(\sqrt{2}\Omega_n\tau)\left[\cos\left(\frac{A\tau}{2}\right)\sigma_x^n + \sin\left(\frac{A\tau}{2}\right)\sigma_z^e\sigma_x^n\right]\end{aligned}$$

where we took the limit $A \gg \Omega_n$ to neglect off-resonance driving. Owing to the combination of the hyperfine coupling $A\sigma_z^e\sigma_z^n/4$ and the radio-frequency driving $\Omega_n B_{1x}\sigma_x^n/2$, the propagators of the nuclear spin before and after the microwave π pulse do not commute and the nuclear spin evolves about non-parallel axes in the two time intervals. This can be directly observed as a modulation of the electronic spin echo, similar to the more common Electron Spin Echo Envelope Modulation (ESEEM)³⁴ for anisotropic hyperfine coupling. To avoid the effects of the hyperfine, we set $A\tau = 2\pi$. Because $A \gg \Omega_n$, we can choose a timing τ that fulfils both the above condition and the desired radio-frequency pulse time, yielding $V(\tau) = \sigma_x^n$. *Measurement model and nuclear-spin-dependent fluorescence intensity.* Measurement of the NV spin state is achieved by monitoring the photoluminescence under laser excitation at 532 nm. Owing to spin-dependent photodynamics in the excited state, the photoluminescence intensity is correlated with the NV electronic spin population. At magnetic fields close to the ESLAC, the photoluminescence intensity becomes modulated by the nuclear spin state as well, owing to the strong hyperfine coupling between the electronic and nuclear spins in the excited state^{21,35}. Therefore, the observable in the experiment can be modelled by the operator

$$M = n_{1,1}|1, 1\rangle\langle 1, 1| + n_{1,0}|1, 0\rangle\langle 1, 0| + n_{0,1}|0, 1\rangle\langle 0, 1| + n_{0,0}|0, 0\rangle\langle 0, 0|$$

where n_{m_q, m_a} are stochastic variables denoting the total number of photons³⁶ detected during the measurement time (300 ns) from the initial state $|m_q, m_a\rangle$. In writing the observable, we neglected the $m_S = +1$ and $m_I = +1$ states, which are never populated in the experiment. To reduce the photon shot noise, all measurements are repeated 5×10^5 times. The raw photoluminescence signal S at each measured point of the sequence is normalized by the photoluminescence collected from two fiducial states $|1, 1\rangle$ and $|0, 1\rangle$. These states are prepared at each average from optical polarization and optical polarization followed by adiabatic passage of the electronic spin to the $|1, 1\rangle$ state. The raw signal from a state ρ , is then normalized to yield

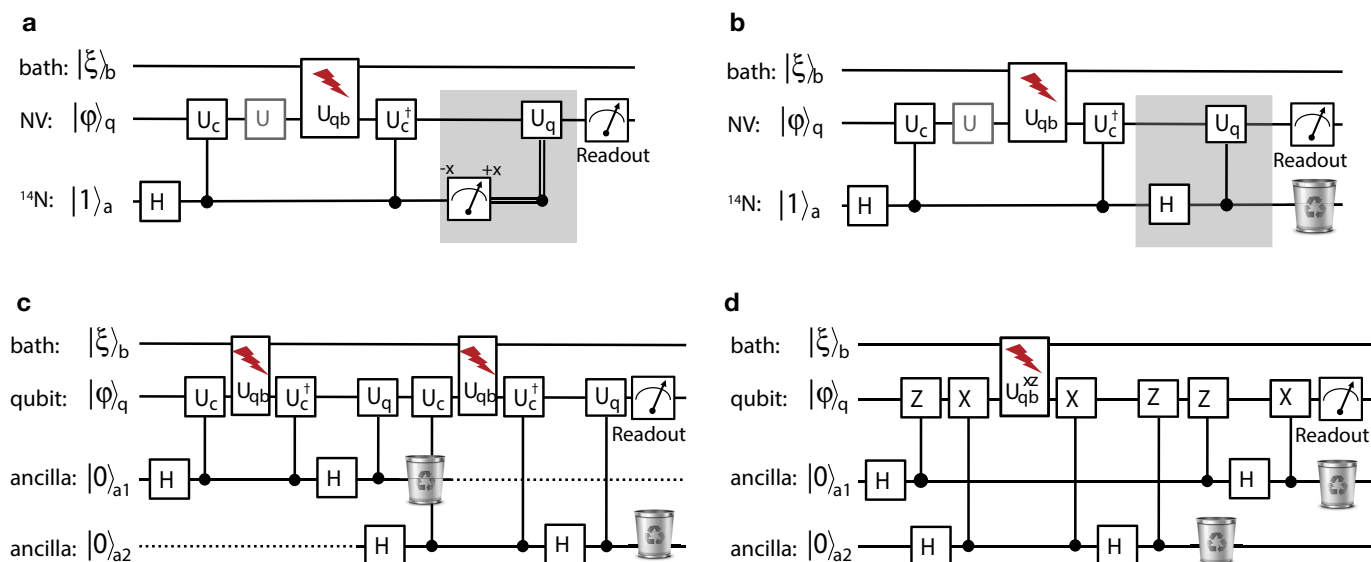
$$\bar{S} = \frac{\text{Tr}(M\rho) - \langle n_{0,1} \rangle}{\langle n_{1,1} \rangle - \langle n_{0,1} \rangle} = \text{Tr}(\bar{M}\rho)$$

where we define the normalized measurement operator \bar{M} as

$$\begin{aligned}\bar{M} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - \epsilon & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta \end{bmatrix} \\ &= \frac{1}{4}[(2 - \eta - \epsilon)\mathbb{1} + (\epsilon + \eta)\sigma_z^n + (\eta - \epsilon + 2)\sigma_z^e + (\epsilon - \eta)\sigma_z^e\sigma_z^n]\end{aligned}$$

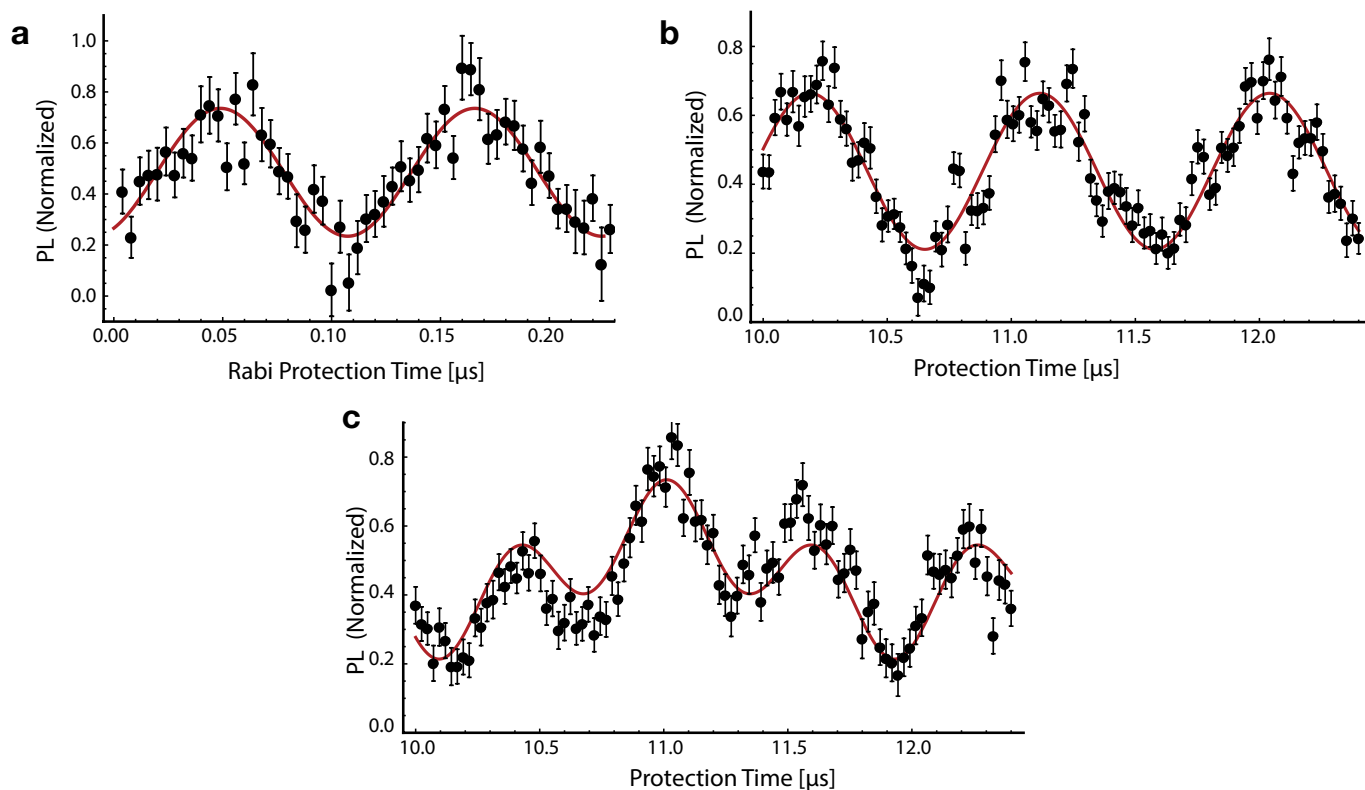
Here $\epsilon = (\langle n_{1,1} \rangle - \langle n_{1,0} \rangle) / (\langle n_{1,1} \rangle - \langle n_{0,1} \rangle)$ and $\eta = (\langle n_{0,1} \rangle - \langle n_{0,0} \rangle) / (\langle n_{1,1} \rangle - \langle n_{0,1} \rangle)$ describe the optical distinguishability of the nuclear spin states in each electronic manifold. Usually, the nuclear spin state does not affect photoluminescence intensity, that is, $\epsilon = \eta = 0$, because it does not affect the relaxation process of the NV centre. However, close to the ESLAC, dynamics driven by the strong excited-state hyperfine coupling renders the relaxation process of the NV spin correlated with the nuclear spin state²⁶, leading to an optical contrast among the nuclear spin states. Therefore, the strength of measurement of the nuclear spin can be varied by changing the static magnetic field. In our experiments we choose two working conditions that, although still allowing nuclear spin polarization ($P > 0.9$) by optical pumping, could highlight differences due to partial measurement of the ancilla spin. Consequently, we performed the experiment at two different magnetic fields ($B = 390$ G and $B = 514$ G) characterized by different strengths of the nuclear-spin optical contrast ($\epsilon = 0.14$, $\eta = 0.16$ at $B = 390$ G and $\epsilon = 0.86$, $\eta = 0.25$ at $B = 514$ G). These values were estimated by measuring the photoluminescence intensity for all four states (see Extended Data Fig. 5).

31. Lindblad, G. On the generators of quantum dynamical semigroups. *Commun. Math. Phys.* **48**, 119–130 (1976).
32. Bloch, F. Generalized theory of relaxation. *Phys. Rev.* **105**, 1206–1222 (1957).
33. Kälin, M. & Schweiger, A. Radio-frequency-driven electron spin echo envelope modulation spectroscopy on spin systems with isotropic hyperfine interactions. *J. Chem. Phys.* **115**, 10863–10875 (2001).
34. Mims, W. B. Envelope modulation in spin-echo experiments. *Phys. Rev. B* **5**, 2409–2419 (1972).
35. Neumann, P. *et al.* Excited-state spectroscopy of single NV defects in diamond using optically detected magnetic resonance. *New J. Phys.* **11**, 013017 (2009).
36. Meriles, C. A. *et al.* Imaging mesoscopic nuclear spin noise with a diamond magnetometer. *J. Chem. Phys.* **133**, 124105 (2010).
37. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* Ch. 4 (Cambridge Univ. Press, 2003).



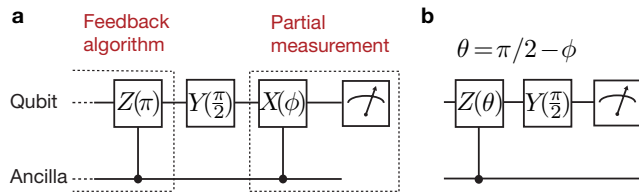
Extended Data Figure 1 | Feedback circuit. See, for example, ref. 37 for an explanation of the notation. **a**, **b**, Measurement-based (**a**) and coherent (**b**) feedback algorithms. In the shaded regions we highlight the differences between the two strategies. The measurement-based protocol requires a measurement of the ancilla and subsequent classically controlled operation (double lines indicate a classical wire). The coherent feedback protocol

does not perform a measurement, but requires a coherent controlled operation. **c**, Re-initializing the ancilla or using multiple fresh ancillas can extend the feedback protection beyond the coherence time of the ancilla. **d**, A concatenated feedback algorithm with two ancillas can protect the qubit from general noise (applied along any axis).

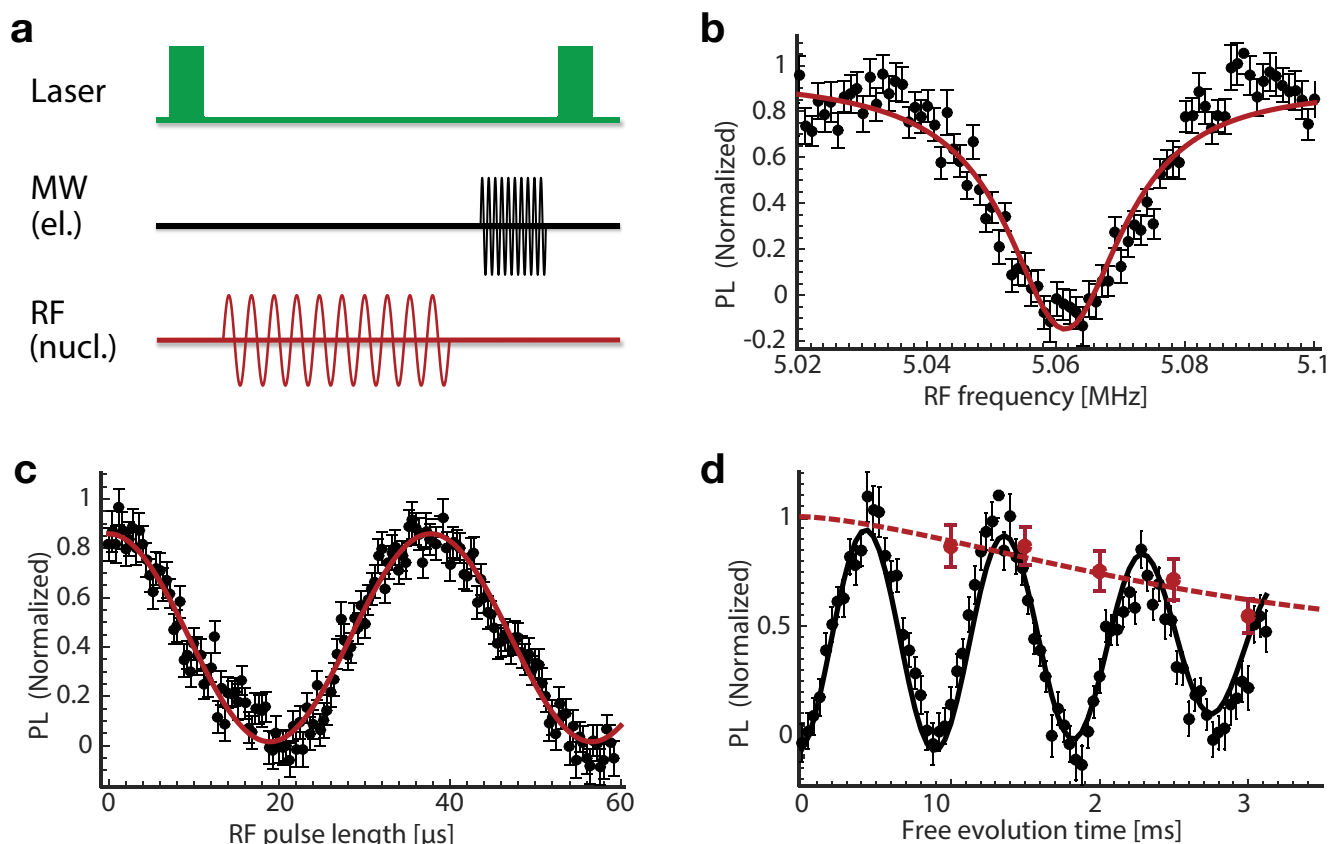


Extended Data Figure 2 | Protected gate. The feedback algorithm is compatible with the application of NOT gates at any point during the protection time. **a**, Rabi oscillations embedded in the feedback-based protection algorithm. **b**, When the NOT gate is applied in the middle of the protection time, it halves the period of the oscillations due to

the hyperfine coupling; see equation (1). **c**, More complex evolution is obtained when inserting the NOT gate at other times. Here we show the behaviour for $\tau_\pi = \tau/4$. Black circles are experimental data with error bars representing their standard deviation (see Fig. 2); the solid lines are fits using equation (1). PL, photoluminescence.

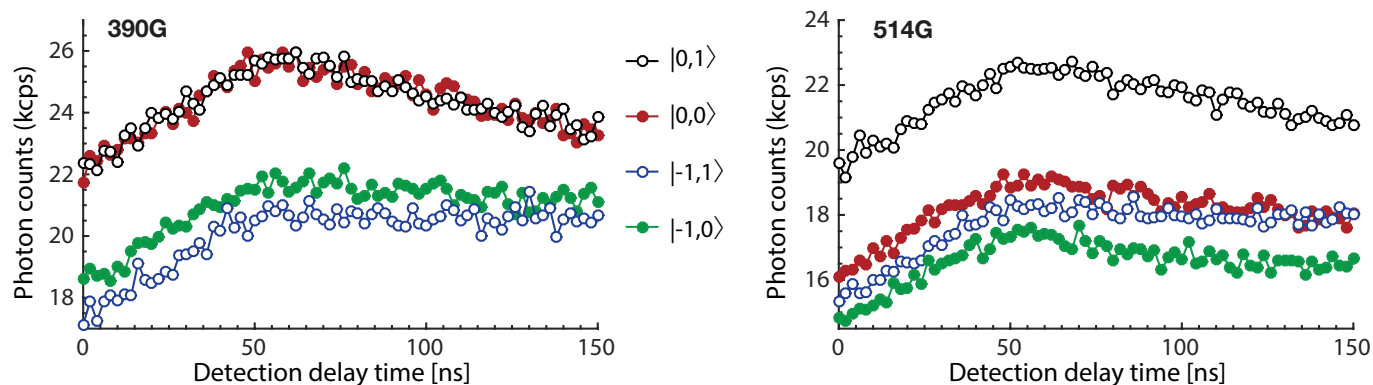


Extended Data Figure 3 | Weak measurement circuit. a, After the feedback algorithm is completed, a controlled- X rotation entangles the qubit with the ancilla, allowing us to perform a partial measurement of the ancilla: $|\psi\rangle_e (\alpha|1\rangle_n + \beta|0\rangle_n) \rightarrow \alpha \cos(\phi)|\psi\rangle_e |1\rangle_n + \beta \sin(\phi)(\sigma_x|\psi\rangle_e)|0\rangle_n$. The strength of the measurement can be adjusted by the angle ϕ . **b**, The control sequence in **a** is simplified by combining the partial measurement gates with the preceding gates of the feedback algorithm. As a result, to implement the weak correlated measurement in the experiment, we simply perform a controlled-phase rotation gate with $\theta = \pi/2 - \phi$, instead of the controlled- Z ($\theta = \pi$) rotation that is required for the feedback algorithm.



Extended Data Figure 4 | Control and coherence of the nitrogen nuclear spin. **a**, Typical control sequence. The first laser pulse initializes the spins into the $|0, 1\rangle$ state. After a nuclear-spin gate (radio-frequency (RF) driving), to detect the nuclear spin state we use a microwave (MW)-selective π pulse that maps the nuclear spin state onto that of the NV spin, which is subsequently detected by the second laser pulse. **b**, Nuclear magnetic resonance at about 390 G. We sweep the radio frequency at fixed pulse length. The dip in the photoluminescence (PL) spectrum indicates the resonant frequency of the $|0, 1\rangle \leftrightarrow |0, 0\rangle$ transition. **c**, Nuclear Rabi oscillations. We drive resonantly the $|0, 1\rangle \leftrightarrow |0, 0\rangle$ transition. The

measured nutation rate (Rabi frequency) is $\Omega_n \approx 26.3$ kHz (ref. 20). **d**, Nuclear Ramsey fringes under the sequence $\pi/2 - \tau - \pi/2$, where τ is the free evolution time. We measure a dephasing time of the nuclear spin $T_{2n}^* = 3.2$ ms, which is limited by the NV electronic spin lattice relaxation process ($T_1 = 4.5$ ms; red circles and dashed line). Error bars represent the standard deviation of the signal, calculated by error propagation from the photoluminescence intensity of the signal and reference photoluminescence curves acquired for each data point for $m_S = \{0, -1\}$. Lines are fits to the data.



Extended Data Figure 5 | Electronic- and nuclear-spin-dependent fluorescence at different magnetic field strengths. At the lower magnetic field (left; $B = 390$ G), the fluorescence intensity shows only a weak dependence on the nuclear spin state in the $m_S = -1$ manifold (indicated by the overlapping data), whereas at $B = 514$ G, which is very close to the ESLAC, a strong dependence on the nuclear spin state is observed in both

manifolds (right; indicated by the non-overlapping data). From these fluorescence measurements, we obtained the parameters (ϵ, η) used to model the measurement operator. In the experiments, the detection time delay and window were optimized to obtain the maximum contrast of the state at each magnetic field.

Quantum hydrogen–bond symmetrization in the superconducting hydrogen sulfide system

Ion Errea^{1,2}, Matteo Calandra³, Chris J. Pickard⁴, Joseph R. Nelson⁵, Richard J. Needs⁵, Yinwei Li⁶, Hanyu Liu⁷, Yunwei Zhang⁸, Yanming Ma⁸ & Francesco Mauri^{3,9}

The quantum nature of the proton can crucially affect the structural and physical properties of hydrogen compounds. For example, in the high-pressure phases^{1,2} of H₂O, quantum proton fluctuations lead to symmetrization of the hydrogen bond and reduce the boundary between asymmetric and symmetric structures in the phase diagram by 30 gigapascals (ref. 3). Here we show that an analogous quantum symmetrization occurs in the recently discovered⁴ sulfur hydride superconductor with a superconducting transition temperature T_c of 203 kelvin at 155 gigapascals—the highest T_c reported for any superconductor so far. Superconductivity occurs via the formation of a compound with chemical formula H₃S (sulfur trihydride) with sulfur atoms arranged on a body-centred cubic lattice^{5–9}. If the hydrogen atoms are treated as classical particles, then for pressures greater than about 175 gigapascals they are predicted to sit exactly halfway between two sulfur atoms in a structure with $Im\bar{3}m$ symmetry. At lower pressures, the hydrogen atoms move to an off-centre position, forming a short H–S covalent bond and a longer H...S hydrogen bond in a structure with $R3m$ symmetry^{5–9}. X-ray diffraction experiments confirm the H₃S stoichiometry and the sulfur lattice sites, but were unable to discriminate between the two phases¹⁰. *Ab initio* density-functional-theory calculations show that quantum nuclear motion lowers the symmetrization pressure by 72 gigapascals for H₃S and by 60 gigapascals for D₃S. Consequently, we predict that the $Im\bar{3}m$ phase dominates the pressure range within which the high T_c was measured. The observed pressure dependence of T_c is accurately reproduced in our calculations for the $Im\bar{3}m$ phase, but not for the $R3m$ phase. Therefore, the quantum nature of the proton fundamentally changes the superconducting phase diagram of H₃S.

The recent discovery of high-temperature superconductivity in compressed hydrogen sulfide⁴ has led to a number of theoretical studies aimed at understanding the phase diagram of the H–S system and the origin of the astonishingly high T_c that was observed^{5–9,12–18}. The overall consensus is that H₂S, the only stable compound formed by hydrogen and sulfur at ambient conditions, is metastable at high pressures and its decomposition gives rise to several H–S compounds. High- T_c superconductivity is believed to occur in a structure with H₃S stoichiometry, and is considered to be conventional in nature, that is, mediated by electron–phonon interactions^{4,5,7,9,12–17}. Alternatives to conventional superconductivity have also been discussed¹⁸. According to structural predictions^{5–9}, H₃S adopts a rhombohedral $R3m$ form between approximately 112 GPa and 175 GPa, and a cubic $Im\bar{3}m$ form at higher pressures. As shown in Fig. 1, the $R3m$ phase is characterized by covalently bonded SH₃ units with a covalent H–S bond of length d_1 . Each of these H atoms is bonded to the next S atom by a hydrogen

H...S bond of length d_2 . By contrast, the $Im\bar{3}m$ phase has full cubic symmetry, with $d_1 = d_2$ so that each H atom resides midway between the two S atoms, as shown in Fig. 1. The $R3m$ structure is nevertheless very close to cubic symmetry; for example, the density functional theory (DFT)-relaxed $R3m$ structure, which represents the minimum of the Born–Oppenheimer energy surface (BOES), has a rhombohedral angle of 109.49° at approximately 150 GPa, compared to 109.47° for a perfect bcc lattice. We verified that imposing a cubic angle on the $R3m$ structure has a negligible effect on the energy difference between the $R3m$ and $Im\bar{3}m$ structures. Consequently, we assume a cubic lattice for both phases in the following.

The bond-symmetrizing second-order transition from $R3m$ to $Im\bar{3}m$ occurs at 175 GPa according to our static lattice calculations. At this pressure, our harmonic phonon calculations show that a Γ -point optical phonon of the high-symmetry $Im\bar{3}m$ phase becomes imaginary, implying that $Im\bar{3}m$ is at a saddle point of the BOES between 112 GPa and 175 GPa, whereas the $R3m$ phase lies at the minimum. Crystal symmetry guarantees that the transition is second-order (see Methods for a symmetry analysis). As occurs in the high-pressure ice X phase^{3,19,20} and other hydrogenated compounds²¹, the quantum nature of the proton can radically alter the pressure at which the second-order phase transition occurs and, in the present case, can strongly affect the stability of the $R3m$ phase below 175 GPa. Determining the stability ranges of these phases therefore requires the inclusion of vibrational zero-point energy (ZPE) along with the static BOES energy. However, the presence of imaginary phonon frequencies

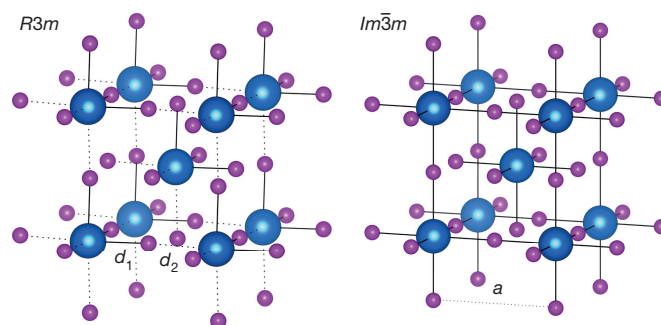


Figure 1 | Crystal structures of the competing phases. Crystal structure in the conventional bcc cell of the $R3m$ (left) and $Im\bar{3}m$ (right) phases. In the $R3m$ structure, the H–S covalent bond of length d_1 is marked with a solid line and the longer H...S hydrogen bond of length d_2 with a dotted line. In the $Im\bar{3}m$ phase, $d_1 = d_2$. The lattice parameter a is marked in the $Im\bar{3}m$ structure. Blue and pink atoms represent S and H atoms, respectively.

¹Fisika Aplikatua 1 Saila, EUITI Bilbao, University of the Basque Country (UPV/EHU), Rafael Moreno “Pitxitxi” Pasealekua 3, 48013 Bilbao, Spain. ²Donostia International Physics Center (DIPC), Manuel Lardizabal Pasealekua 4, 20018 Donostia/San Sebastián, Spain. ³IMPMC, UMR CNRS 7590, Sorbonne Universités – UPMC Université Paris 06, MNHN, IRD, 4 Place Jussieu, F-75005 Paris, France. ⁴Department of Materials Science and Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, UK. ⁵Theory of Condensed Matter Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK. ⁶School of Physics and Electronic Engineering, Jiangsu Normal University, Xuzhou 221116, China. ⁷Geophysical Laboratory, Carnegie Institution of Washington, Washington DC 20015, USA. ⁸State Key Laboratory of Superhard Materials, Jilin University, Changchun 130012, China. ⁹Dipartimento di Fisica, Università di Roma “La Sapienza”, Piazzale Aldo Moro 5, I-00185 Roma, Italy.

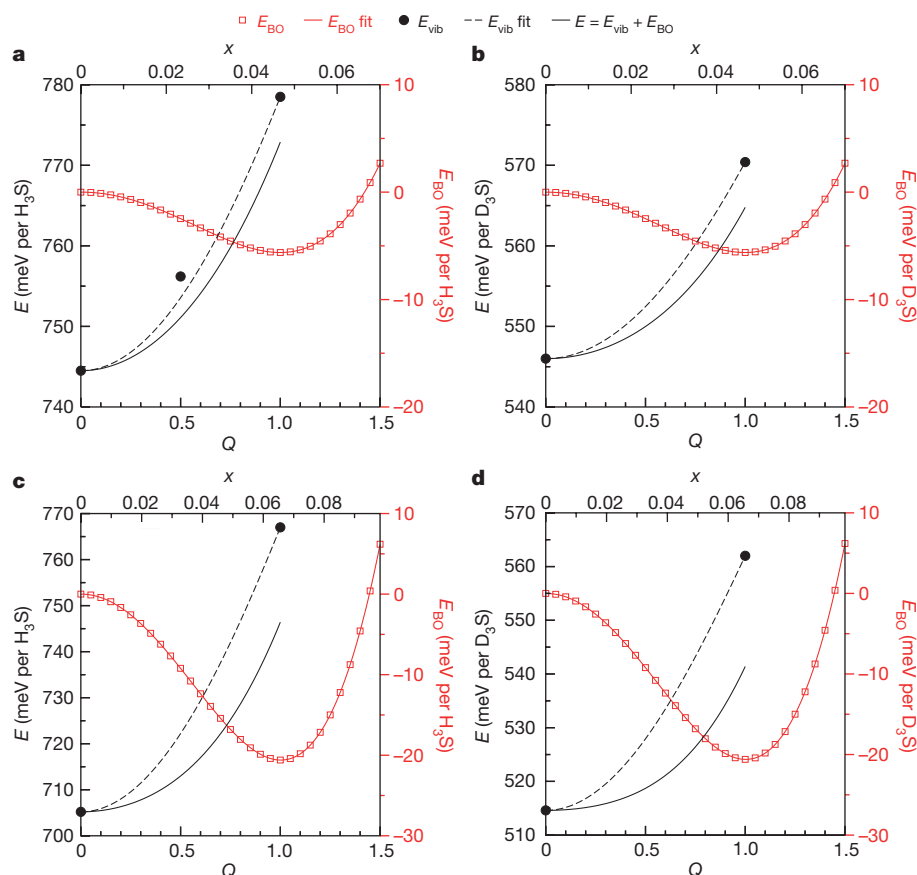


Figure 2 | Energetics. **a–d**, E_{BO} , E_{vib} and $E = E_{\text{vib}} + E_{\text{BO}}$ curves for H_3S (**a**, **c**) and D_3S (**b**, **d**) are shown as functions of the reaction coordinate Q that transforms the $Im\bar{3}m$ structure ($Q = 0$) into the $R3m$ structure ($Q = 1$), as well as the relative coordinate $x = (d_2 - a/2)/(a/2)$ that measures the off-centring of the H atoms, where d_2 is the length of the hydrogen bond and a is the lattice parameter (see Fig. 1). The left-hand (black) axes show the energy scale for E_{vib} and E ; the right-hand (red) axes show the scale for E_{BO} . Crystal symmetry implies that $E(Q) = E(-Q)$, so that the curves can be fitted to polynomials with only even terms; this guarantees that the

transition is second-order according to Landau theory¹¹. Results are presented for two different volumes of the primitive bcc lattice; $V = 97.85a_0^3$ (**a**, **b**) corresponds to approximately 150 GPa and $V = 102.11a_0^3$ (**c**, **d**) to 130 GPa. The pressure associated with each volume depends on both the isotope and Q (see Extended Data Fig. 1 for the equations of state). Black circles represent calculated E_{vib} points and the black dashed line the fitted $E_{\text{vib}}(Q)$ curve (see Methods). The $E(Q)$ curve is obtained by addition of the fitted E_{vib} and E_{BO} curves.

hinders calculations of the ZPE, because the quasi-harmonic approximation breaks down, and anharmonicity becomes important.

To elucidate the role of anharmonicity and quantum effects in the pressure range in which the record $T_c = 203$ K was observed, we make use of the stochastic self-consistent harmonic approximation (SSCHA)^{22,23}. The variational SSCHA method was devised for calculating the free energy and phonon spectra while fully incorporating quantum and anharmonic effects, and so is perfectly suited for our purpose. All of the calculations presented here are performed at 0 K. Primitive cells for the $R3m$ and $Im\bar{3}m$ structures contain four atoms (one S atom and three H atoms); consequently, a particular nuclear configuration can be described by a 12-dimensional vector \mathbf{R} containing the atomic coordinates. In the classical limit, the ZPE is neglected and the energy of a nuclear configuration \mathbf{R} is given by the DFT Born–Oppenheimer energy $E_{\text{BO}}(\mathbf{R})$. In the SSCHA, the ZPE is accounted for by approximating the nuclear wavefunction by a Gaussian centred on a centroid coordinate \mathbf{R}_c , which denotes the average and most probable position of the nuclei. For a given \mathbf{R}_c , the width of the Gaussian is obtained by variational minimization of the expectation value of the sum of the nuclear potential and kinetic energies. In the following analysis, it is convenient to split the SSCHA total energy $E(\mathbf{R}_c)$ into static and anharmonic-vibrational-ZPE contributions: $E(\mathbf{R}_c) = E_{\text{BO}}(\mathbf{R}_c) + E_{\text{vib}}(\mathbf{R}_c)$.

We study the energy landscape $E(\mathbf{R}_c)$ along the line defined by $\mathbf{R}_c(Q) = \mathbf{R}_{Im\bar{3}m} + Q(\mathbf{R}_{R3m} - \mathbf{R}_{Im\bar{3}m})$, in which $\mathbf{R}_{Im\bar{3}m}$ and \mathbf{R}_{R3m} are,

respectively, the coordinates corresponding to the saddle point and minimum of the BOES, representing the two different symmetries. Here, Q is a real number describing the reaction coordinate, so that at $Q = 0$ the centroids are located at the atomic positions of $Im\bar{3}m$, and at $Q = 1$ at the atomic positions of $R3m$. Hence, Q measures the off-centring of the hydrogen nuclear wavefunction and can be associated with the relative coordinate $x = (d_2 - a/2)/(a/2)$ that quantifies the length of the $\text{H}\cdots\text{S}$ hydrogen bond with respect to the symmetric position (a is the lattice parameter). We analyse the curve $E(\mathbf{R}_c(Q)) \equiv E(Q)$ for a fixed primitive bcc unit cell. As shown in Fig. 2a for a cell volume of $97.85a_0^3$ (a_0 is the Bohr radius), the $E_{\text{BO}}(Q)$ curve has a shallow double-well structure that favours the $R3m$ structure by only 5.6 meV per H_3S . However, after adding the $E_{\text{vib}}(Q)$ energy calculated with the SSCHA, the full $E(Q)$ curve shows a clear minimum at $Q = 0$, which favours the $Im\bar{3}m$ structure. At a larger volume of $102.11a_0^3$, which corresponds to a pressure of about 130 GPa, the minimum of $E(Q)$ is also at $Q = 0$, despite the fact that the one-dimensional Born–Oppenheimer well in $E_{\text{BO}}(Q)$ becomes deeper, as shown in Fig. 2c. Upon repeating these calculations for D_3S , we find that the $Im\bar{3}m$ structure is the most favourable once the ZPE has been included. Therefore, we conclude that the quantum nature of the nuclei symmetrizes the hydrogen bond and leads to a proton wavefunction that is centred at the atomic positions of $Im\bar{3}m$ for both H_3S and D_3S . To eliminate the possibility that the energy minimum occurs beyond the $\mathbf{R}_c(Q)$ line studied, we performed an unconstrained SSCHA

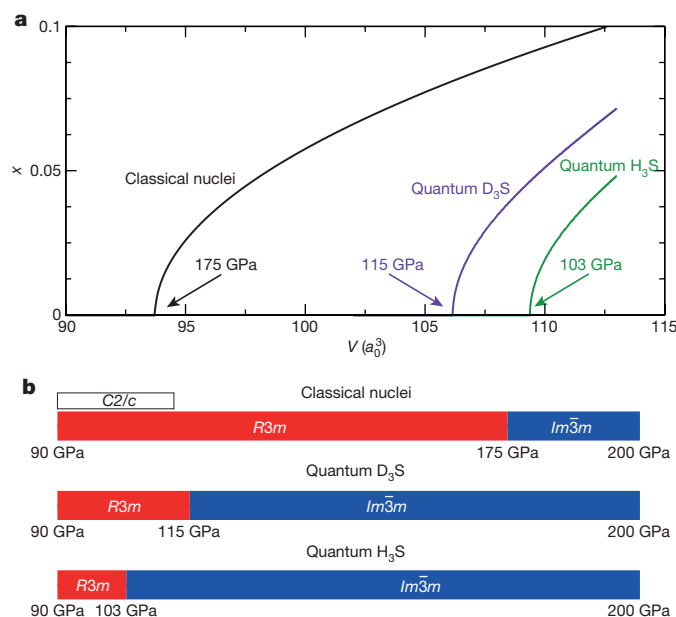


Figure 3 | Second-order phase transition. **a**, For each volume V we plot the relative coordinate x that yields the minimum total energy; x measures the off-centring of the H atoms (see Fig. 2). The results are shown in the classical nuclei limit (black), as well as in the quantum case both for H_3S (green) and D_3S (purple). The volume at which x departs from zero marks the second-order phase transition from the $Im\bar{3}m$ phase to the $R3m$ phase. Transition pressures are also indicated, and include the effects of vibrational energies. **b**, Phase diagram for the second-order phase transition as a function of pressure. As shown in ref. 8, below 112 GPa, H_3S adopts a very different $C2/c$ phase; we mark the expected emergence of this phase by a box.

minimization, optimizing both the width of the Gaussians and the R_c centroid positions. The results of this minimization confirm that, within stochastic error, the centroid position obtained corresponds to

the $Im\bar{3}m$ structure, in which the H–S covalent and H...S hydrogen bond distances equalize, leading to symmetric hydrogen bonds.

The difference between the vibrational energies of $R3m$ and $Im\bar{3}m$ as a function of the x coordinate is weakly dependent on volume. This allows us to interpolate $E(x)$ in a wide volume range and estimate the pressure at which the proton wavefunction shifts away from the centred position. Our calculations show that this symmetry breaking occurs at 103 GPa in H_3S and 115 GPa in D_3S (see Fig. 3). The higher transition pressure in D_3S is due to weaker quantum effects. This isotope effect is similar to that observed in the ice VII/ice X transition²⁴. Considering that, below 112 GPa, the $R3m$ phase is expected to transform into a very different monoclinic $C2/c$ phase consisting of isolated H_2S and H_2 molecules with H_3S stoichiometry⁸, $R3m$ H_3S might not be formed. However, D_3S may adopt the $R3m$ structure at pressures below the transition to the $Im\bar{3}m$ phase.

The quantum proton symmetrization has a very large impact on the phonon spectra of H_3S . As mentioned previously, and shown in Fig. 4, the phonon spectra of $Im\bar{3}m$ H_3S have several imaginary modes in the harmonic approximation below 175 GPa. The corresponding anharmonic SSCHA phonon spectra for $Im\bar{3}m$ H_3S show well-behaved phonon dispersion relations with positive frequencies in the pressure range of interest (Fig. 4). This behaviour is analogous to ice X, which has only real positive phonon frequencies in the regime in which the classical limit predicts symmetrization of the hydrogen bond^{25–28}. The anharmonic renormalization of the phonon energies is huge, especially for the H–S bond-stretching modes in which H atoms move towards the neighbouring S atoms, which are precisely those modes that drive the second-order phase transition between the $Im\bar{3}m$ and $R3m$ phases. Therefore, the proximity to the second-order quantum phase transition is the origin of the strong anharmonicity.

Although the bond symmetrization in ice X occurs in an insulating system, H_3S is metallic and the symmetrization strongly affects the superconductivity. The calculated electron–phonon coupling and T_c lend further support to the suggestion that $Im\bar{3}m$ H_3S yields the record $T_c = 203$ K. We use Wannier-interpolated electron–phonon matrix

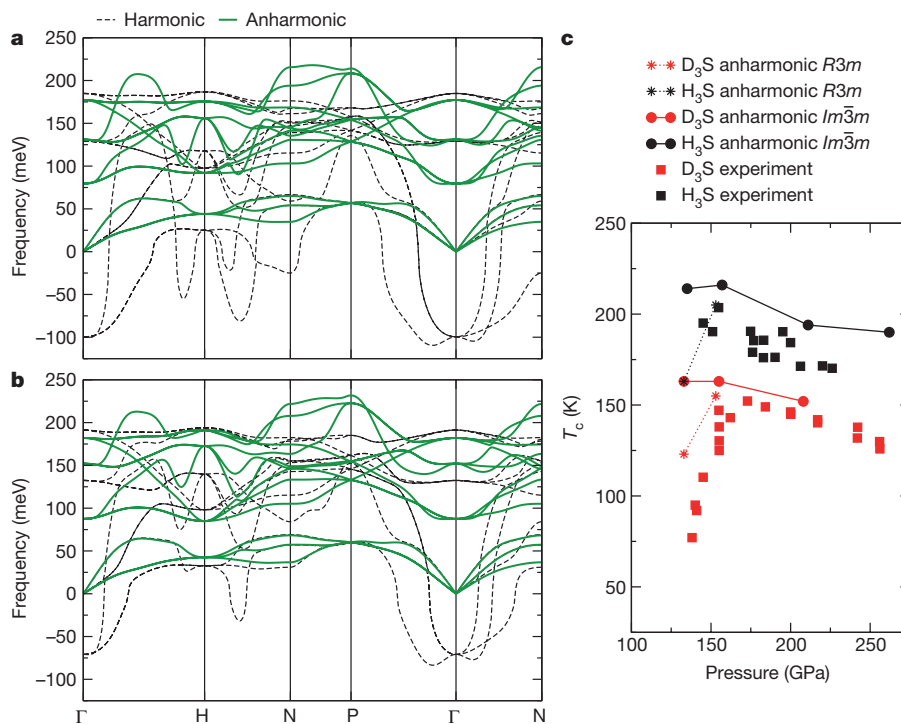


Figure 4 | Phonon spectra and superconducting transition temperature. **a**, **b**, Harmonic and SSCHA anharmonic phonon spectra of the cubic high-symmetry $Im\bar{3}m$ structure for H_3S at different pressures: 135 GPa (**a**) and 157 GPa (**b**). **c**, T_c calculated with the anharmonic phonons for the $Im\bar{3}m$

structure compared with experimental results obtained after annealing⁴. T_c results obtained with anharmonic phonons for the $R3m$ structure below 175 GPa are also shown. Each pressure takes the vibrational energy into account.

elements in our calculations²⁹ and estimate T_c by solving the isotropic Migdal–Eliashberg equations. The phonon frequencies and polarizations that enter the electron–phonon calculations are calculated using the SSCHA. Therefore, in the present treatment we neglect nonlinear corrections of the electron–phonon vertices and non-adiabatic effects that could arise from the small mass of H. The results obtained for the $Im\bar{3}m$ structure using anharmonic phonon frequencies agree well with experimental measurements of T_c for H_2S and D_2S , and correctly capture the observed increase in T_c with decreasing pressure. We also find an isotope coefficient $\alpha = -\{\ln[T_c(D_2S)] - \ln[T_c(H_2S)]\}/\ln(2)$ for $H \rightarrow D$ substitution of $\alpha = 0.35$ at 210 GPa and $\alpha = 0.40$ at 155 GPa in good agreement with experiment (see Fig. 4c). The electron–phonon coupling constant λ , which scales with the phonon frequencies as $1/\omega^2$, is enhanced with decreasing pressure owing to the overall softening of the phonon modes. This explains the smooth decrease in T_c with increasing pressure. Between approximately 130 GPa and 150 GPa, the increase in λ is compensated by the decrease in the average phonon frequency, and T_c saturates. We also present SSCHA calculations for the $R\bar{3}m$ structure keeping the centroids at the $Q = 1$ position. We find a rapid decrease in T_c with decreasing pressure as in previous harmonic calculations¹³. Therefore, the observed high- T_c superconductivity cannot be explained by H_2S in the $R\bar{3}m$ phase, although the sudden decrease in T_c measured for D_2S below 150 GPa (ref. 4) could be an indication of the symmetry breaking that we predicted at 115 GPa. Indeed, the predicted transition pressure depends on the choice of the exchange–correlation functional (see Methods). Even if our choice of the PBE (Perdew–Burke–Ernzerhof) exchange–correlation functional³⁰ seems appropriate, on the basis of the agreement between the experimentally observed equation of state¹⁰ and DFT calculations, we cannot exclude a small error in the transition pressure. However, the isotopic shift in the transition pressure is independent of the functional. Thus, if the decrease in T_c observed experimentally in D_2S is due to the transition from $R\bar{3}m$ to $Im\bar{3}m$, then we predict that a similar decrease also occurs in H_2S , but is shifted to lower pressures by about 12 GPa. Considering that the $Im\bar{3}m$ phase has no Raman active modes, if the drop in T_c in D_2S coincides with the second-order phase transition, then we predict that Raman peaks from the $R\bar{3}m$ phase would emerge at that pressure.

A recent experiment⁴ suggests that room-temperature superconductivity is reachable in other hydrogen-rich compounds. Our results show that in such hydrogen-rich materials with prospects for high- T_c superconductivity, the quantum motion of the proton induces non-trivial effects that strongly affect the thermodynamical stability, the hydrogen chemical bonding and the electron–phonon coupling strength.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 December 2015; accepted 20 January 2016.

Published online 28 March 2016.

- Goncharov, A. F., Struzhkin, V. V., Somayazulu, M. S., Hemley, R. J. & Mao, H. K. Compression of ice to 210 gigapascals: infrared evidence for a symmetric hydrogen-bonded phase. *Science* **273**, 218–220 (1996).
- Loubeyre, P., LeToullec, R., Wolanin, E., Hanfland, M. & Hausermann, D. Modulated phases and proton centring in ice observed by X-ray diffraction up to 170 GPa. *Nature* **397**, 503–506 (1999).
- Benoit, M., Marx, D. & Parrinello, M. Tunneling and zero-point motion in high-pressure ice. *Nature* **392**, 258–261 (1998).
- Drozdov, A. P., Erements, M. I., Troyan, I. A., Ksenofontov, V. & Shylin, S. I. Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system. *Nature* **525**, 73–76 (2015).
- Duan, D. *et al.* Pressure-induced metallization of dense $(H_2S)_2H_2$ with high- T_c superconductivity. *Sci. Rep.* **4**, 6968 (2014).
- Duan, D. *et al.* Pressure-induced decomposition of solid hydrogen sulfide. *Phys. Rev. B* **91**, 180502 (2015).
- Errea, I. *et al.* High-pressure hydrogen sulfide from first principles: a strongly anharmonic phonon-mediated superconductor. *Phys. Rev. Lett.* **114**, 157004 (2015).

- Li, Y. *et al.* Dissociation products and structures of solid H_2S at strong compression. *Phys. Rev. B* **93**, 020103(R) (2016).
- Bernstein, N., Hellberg, C. S., Johannes, M. D., Mazin, I. I. & Mehl, M. J. What superconducts in sulfur hydrides under pressure and why. *Phys. Rev. B* **91**, 060511 (2015).
- Einaga, M. *et al.* Crystal structure of 200 K-superconducting phase of sulfur hydride system. Preprint at <http://arXiv.org/abs/1509.03156> (2015).
- Landau, L. D. & Lifshitz, E. M. *Statistical Physics* 3rd edn, Vol. 5 of *Course of Theoretical Physics* (Butterworth-Heinemann, 1980).
- Papaconstantopoulos, D. A., Klein, B. M., Mehl, M. J. & Pickett, W. E. Cubic H_2S around 200 GPa: an atomic hydrogen superconductor stabilized by sulfur. *Phys. Rev. B* **91**, 184511 (2015).
- Akashi, R., Kawamura, M., Tsuneyuki, S., Nomura, Y. & Arita, R. First-principles study of the pressure and crystal-structure dependences of the superconducting transition temperature in compressed sulfur hydrides. *Phys. Rev. B* **91**, 224513 (2015).
- Nicol, E. J. & Carbotte, J. P. Comparison of pressurized sulfur hydride with conventional superconductors. *Phys. Rev. B* **91**, 220507 (2015).
- Flores-Livas, J. A., Sanna, A. & Gross, E. K. U. High temperature superconductivity in sulfur and selenium hydrides at high pressure. Preprint at <http://arXiv.org/abs/1501.06336> (2015).
- Li, Y., Hao, J., Liu, H., Li, Y. & Ma, Y. The metallization and superconductivity of dense hydrogen sulfide. *J. Chem. Phys.* **140**, 174712 (2014).
- Bianconi, A. & Jarlborg, T. Superconductivity above the lowest earth temperature in pressurized sulfur hydride. *Europhys. Lett.* **112**, 37001 (2015).
- Hirsch, J. E. & Marsiglio, F. Hole superconductivity in H_2S and other sulfides under high pressure. *Physica C* **511**, 45–49 (2015).
- Lee, C., Vanderbilt, D., Laasonen, K., Car, R. & Parrinello, M. *Ab initio* studies on high pressure phases of ice. *Phys. Rev. Lett.* **69**, 462–465 (1992).
- Lee, C., Vanderbilt, D., Laasonen, K., Car, R. & Parrinello, M. *Ab initio* studies on the structural and dynamical properties of ice. *Phys. Rev. B* **47**, 4863–4872 (1993).
- McMahon, M. I. *et al.* Geometric effects of deuteration on hydrogen-ordering phase transitions. *Nature* **348**, 317–319 (1990).
- Errea, I., Calandra, M. & Mauri, F. First-principles theory of anharmonicity and the inverse isotope effect in superconducting palladium-hydride compounds. *Phys. Rev. Lett.* **111**, 177002 (2013).
- Errea, I., Calandra, M. & Mauri, F. Anharmonic free energies and phonon dispersions from the stochastic self-consistent harmonic approximation: application to platinum and palladium hydrides. *Phys. Rev. B* **89**, 064302 (2014).
- Song, M., Yamawaki, H., Fujihisa, H., Sakashita, M. & Aoki, K. Infrared investigation on ice VIII and the phase diagram of dense ices. *Phys. Rev. B* **68**, 014106 (2003).
- Goncharov, A. F., Struzhkin, V. V., Mao, H. K. & Hemley, R. J. Raman spectroscopy of dense H_2O and the transition to symmetric hydrogen bonds. *Phys. Rev. Lett.* **83**, 1998–2001 (1999).
- Caracas, R. Dynamical instabilities of ice X. *Phys. Rev. Lett.* **101**, 085502 (2008).
- Marqués, M., Ackland, G. J. & Loveday, J. S. Nature and stability of ice X. *High Press. Res.* **29**, 208–211 (2009).
- Bronstein, Y., Depondt, P., Finocchi, F. & Saitta, A. M. Quantum-driven phase transition in ice described via an efficient Langevin approach. *Phys. Rev. B* **89**, 214101 (2014).
- Calandra, M., Profeta, G. & Mauri, F. Adiabatic and nonadiabatic phonon dispersion in a Wannier function approach. *Phys. Rev. B* **82**, 165111 (2010).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

Acknowledgements We acknowledge financial support from the Spanish Ministry of Economy and Competitiveness (FIS2013-48286-C2-2-P), the EPSRC (UK) (grant numbers EP/J017639/1 and EP/K014560/1), the Cambridge Commonwealth Trust, the National Natural Science Foundation of China (grant numbers 11204111, 11404148, 11274136 and 11534003), the 2012 Changjiang Scholars Program of China, and the Natural Science Foundation of Jiangsu province (grant number BK20130223). C.J.P. acknowledges support from the Royal Society through a Wolfson Research Merit award. Work at Carnegie was supported by EFree, an Energy Frontier Research Center funded by the DOE, Office of Science, Basic Energy Sciences under award number DE-SC-0001057. Computer facilities were provided by PRACE and the Donostia International Physics Center (DIPC).

Author Contributions I.E., M.C. and F.M. performed the anharmonic and superconducting calculations. All authors contributed to the design of the research project and to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.E. (ion.errea@ehu.eus), M.C. (matteo.calandra@impmc.upmc.fr) or F.M. (francesco.mauri@uniroma1.it).

METHODS

Calculation details. Supercell calculations for the SSCHA^{22,23} and linear response calculations³¹ were performed within DFT and the generalized gradient approximation (GGA) functional³⁰ as implemented in the Quantum ESPRESSO³² code. We used ultrasoft pseudopotentials³³, a plane-wave cut-off energy of 60 Ry for the kinetic energy and 600 Ry for the charge density. The charge density and dynamical matrices were calculated using a $32 \times 32 \times 32$ Monkhorst–Pack-shifted electron–momentum grid for the unit cell calculations. This mesh was adjusted accordingly in the supercell calculations. The electron–phonon coupling was calculated by using electron and phonon momentum grids composed of up to $42 \times 42 \times 42$ randomly displaced points in the Brillouin zone. The isotropic Migdal–Eliashberg equations were solved using 512 Matsubara frequencies and a Coulomb pseudopotential $\mu^* = 0.16$.

The SSCHA calculations were performed using a $3 \times 3 \times 3$ supercell for both H_3S and D_3S in the $Im\bar{3}m$ phase, yielding dynamical matrices on a commensurate $3 \times 3 \times 3$ q -point grid. The difference between the harmonic and anharmonic dynamical matrices in the $3 \times 3 \times 3$ phonon momentum grid was interpolated to a $6 \times 6 \times 6$ grid. Upon adding the harmonic matrices to the result, the anharmonic dynamical matrices were obtained on a $6 \times 6 \times 6$ grid. These dynamical matrices were used for the anharmonic electron–phonon coupling calculation. The SSCHA calculations for $Q=1$ were performed with a $2 \times 2 \times 2$ supercell. For consistency, the vibrational energies presented in Fig. 2 were also calculated using a $2 \times 2 \times 2$ supercell. However, the electron–phonon calculations for $Q=1$ were performed with the SSCHA dynamical matrices interpolated to a $6 \times 6 \times 6$ grid from the $2 \times 2 \times 2$ mesh.

The $E_{\text{vib}}(Q)$ curves in Fig. 2 were obtained as follows. E_{vib} was calculated for $Q=0$ and $Q=1$ with the SSCHA. With the SSCHA calculation at $Q=1$, we extracted $\frac{dE_{\text{vib}}}{dQ}(Q=1)$ with no further computational effort. Considering that the derivative of the curve at $Q=0$ vanishes by symmetry, we obtain straightforwardly a potential of the form $E_{\text{vib}}(Q) = A + BQ^2 + CQ^4$ (in which A , B and C are constants). The E_{vib} fit curves presented in Fig. 2 were obtained in this way. The extra point obtained at $Q=0.5$ for H_3S at $V = 97.85a_0^3$ (see Fig. 2a) confirmed the validity of the fitting procedure. The $E_{\text{BO}}(Q)$ BOES energies were calculated for many Q points, yielding an accurate fitting curve. Figure 3 was obtained using a polynomial interpolation of the BOES in the volume range shown and adding the $E_{\text{vib}}^{R3m}(x) - E_{\text{vib}}^{Im\bar{3}m}$ curves that are practically independent of volume.

Symmetry analysis of the second-order phase transition. The bond-symmetrization transition from $R3m$ to $Im\bar{3}m$ is a second-order transition and is driven by the softening of an optical mode at Γ . The mode driving the transition belongs to the irreducible representation T_{1u} , also denoted as Γ_4^- , whose dimension is three (refs 34, 35). This irreducible representation is compatible with a group–subgroup relation between the $Im\bar{3}m$ and $R3m$ space groups. Because the mode driving the transition is at the Γ point, the transition occurs without increasing the size of the unit cell, which contains four atoms in the primitive cell of the bcc lattice of the $Im\bar{3}m$ structure as well as in the rhombohedral lattice of the $R3m$ phase.

In a transition from the $m\bar{3}m(O_h)$ point group to the $3m(C_{3v})$, the most general free energy expansion contains only even terms of the order parameters³⁶. Therefore, according to Landau theory¹¹, the transition must be second-order. Our reaction coordinate Q used to describe the second-order transition is a simplified order parameter. The fact that $E(Q) = E(-Q)$ confirms that the transition is second-order.

Equations of state. In Extended Data Fig. 1 we present the equation of state both for the $Im\bar{3}m$ and $R3m$ phases. The pressure is calculated both with and without the vibrational contributions to the energies of H_3S and D_3S . The curves that include the vibrational contribution to the energy are used to determine the vibrational effects on the calculated pressures. The vibrational energy calculated includes anharmonicity in all cases as calculated by the SSCHA. The vibrational energy has a small contribution to the calculated pressure, which varies slightly with the isotopic mass. The pressure–volume curve can be efficiently fitted to the third-order Birch–Murnaghan equation. The parameters of the fit are given in Extended Data Table 1.

Dependence on the exchange–correlation functional. All calculations presented here are calculated within the GGA under the PBE parameterization³⁰. However, we performed additional calculations within the local density approximation (LDA)³⁷ and the BLYP (Becke–Lee–Yang–Parr)^{38,39} parameterization of the GGA. The second-order phase transition at the static level from $Im\bar{3}m$ to $R3m$, which

occurs at 175 GPa within PBE, is reduced to 145 GPa in the LDA and increased to 257 GPa with the BLYP parameterization. In Extended Data Fig. 2 we compare the relative coordinate $x = (d_2 - a/2)/(a/2)$ as a function of volume, where d_2 is the length of the hydrogen bond and a is the lattice parameter. The x coordinate is obtained both with the static BOES energy E_{BO} and with the total energy $E = E_{\text{BO}} + E_{\text{vib}}$. However, the vibrational energy is assumed to be independent of the functional, and the value obtained with PBE is used for estimating E_{vib} for both LDA and BLYP. We believe that this assumption is justified because the differences between the vibrational energies of $Im\bar{3}m$ and $R3m$ calculated within the SSCHA and using PBE are very weakly volume dependent, as shown in Extended Data Fig. 3. At the static level, the LDA symmetrizes the structure at a pressure 30 GPa below that predicted by PBE, whereas BLYP symmetrizes it at a pressure 82 GPa above that predicted by PBE. When including the vibrational contribution as described above, the transition between $Im\bar{3}m$ and $R3m$ occurs at 170 GPa for H_3S and 185 GPa for D_3S within BLYP. The shift in the transition pressure induced by the ZPE and its isotope dependence is similar for PBE and BLYP. The fully symmetric $Im\bar{3}m$ H_3S structure is always favoured within the LDA for H_3S over the pressure range studied, and it only becomes unfavourable below 89 GPa for D_3S .

Although the choice of exchange–correlation functional adds some uncertainty to the predicted transition pressure, we believe PBE is the most appropriate choice. This statement is based on the fact that PBE best reproduces the experimental equation of state results of ref. 10. BLYP overestimates the volume, especially when the vibrational contribution is considered (see Extended Data Fig. 4). Indeed, it is well known that BLYP vastly overestimates the equilibrium volumes in metals⁴⁰ and, therefore, it is not the best choice for H_3S . On the contrary, when including the vibrational contribution to the pressure, which is calculated in all cases using the PBE functional, the PBE equation of state is in rather good agreement with experimental results¹⁰ and certainly performs better than LDA and BLYP.

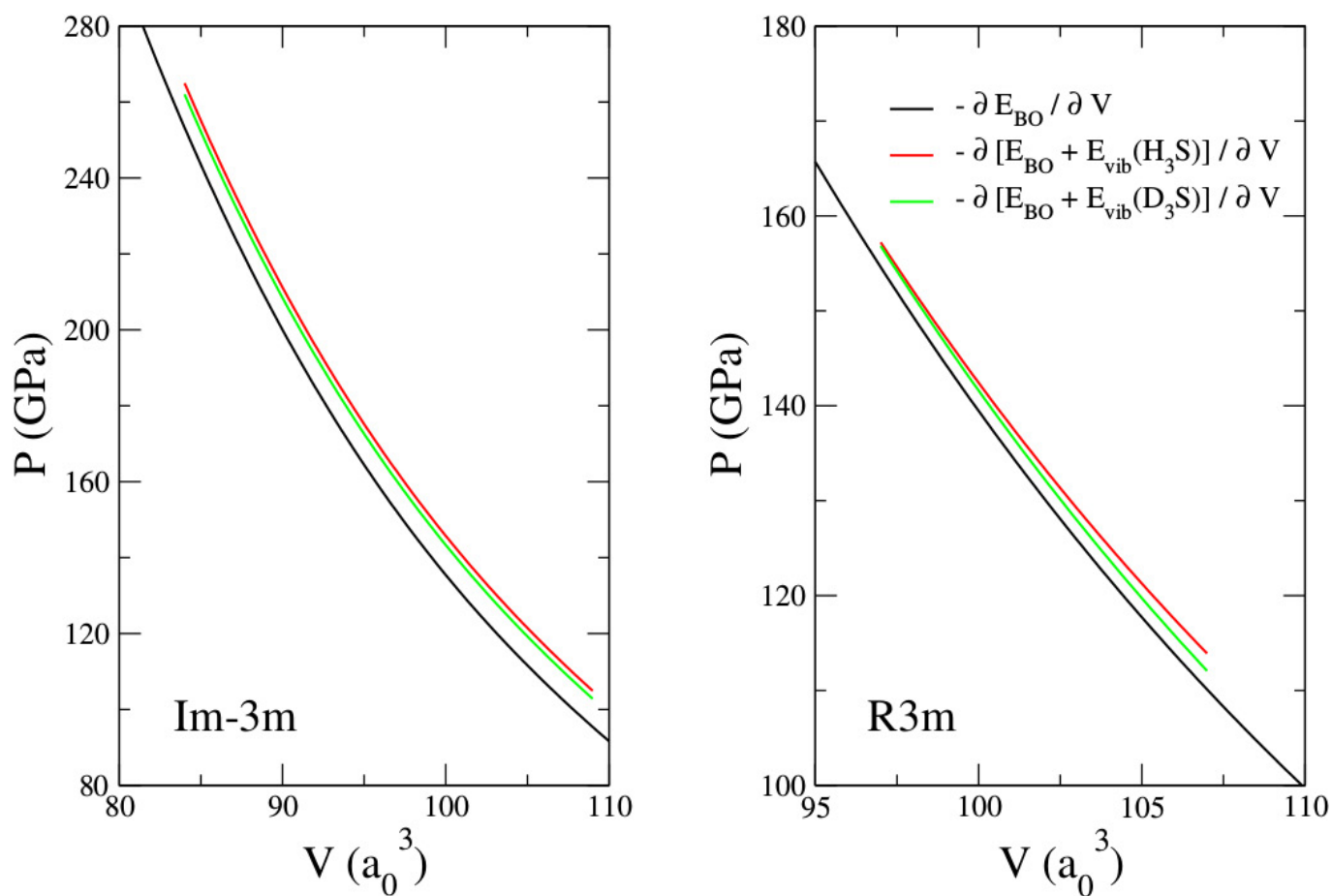
Phonons of $Im\bar{3}m$ D_3S . In Extended Data Fig. 5 we show the harmonic and anharmonic phonon spectra of D_3S in the $Im\bar{3}m$ phase at two different pressures.

Anharmonic phonon dispersion in the $R3m$ phase. In Extended Data Fig. 6 we show the SSCHA phonon spectra calculated for the $R3m$ phase at two different pressures. Here the centroid position of the SSCHA is placed at the minimum of the BOES, which is not a minimum of the total energy, as discussed. The Raman active modes of the $R3m$ phase are listed in Extended Data Table 2.

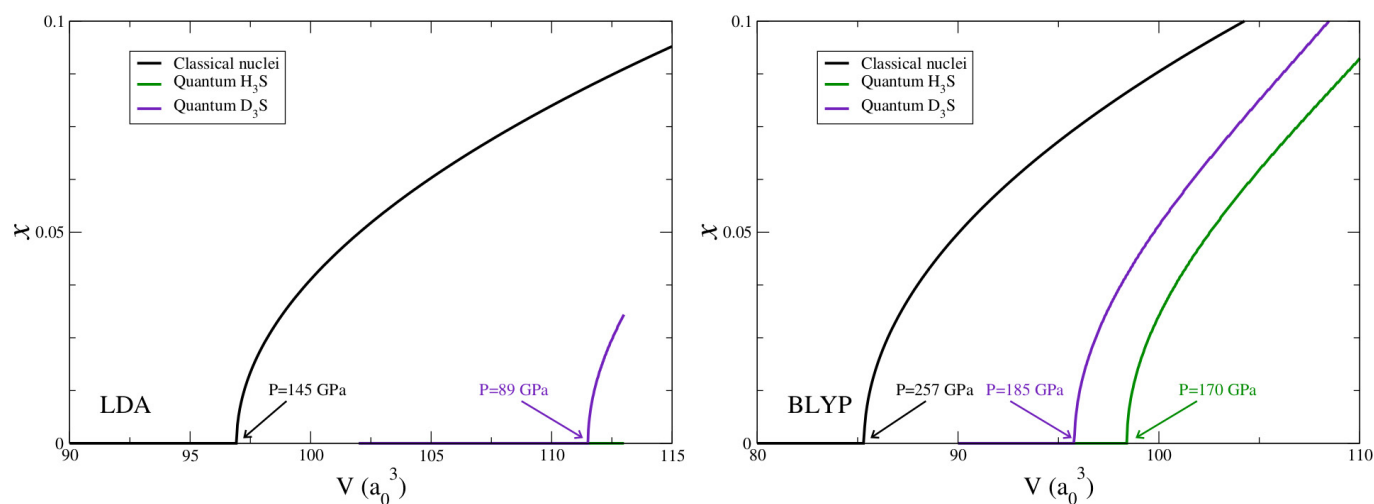
Superconducting properties. In Extended Data Table 3 we summarize the calculated values of the electron–phonon coupling constant λ , logarithmic frequency average ω_{log} and T_c .

In Extended Data Fig. 7 we show the Eliashberg function $\alpha^2F(\omega)$ and integrated electron–phonon coupling constant $\lambda(\omega)$ of the $Im\bar{3}m$ phase at two different pressures calculated using the SSCHA method. The Eliashberg function $\alpha^2F(\omega)$ in the anharmonic case for the $R3m$ structure is given in Extended Data Fig. 6.

- Baroni, S., de Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515–562 (2001).
- Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
- Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41**, 7892–7895 (1990).
- Aroyo, M. I. et al. Bilbao Crystallographic Server: I. Databases and crystallographic computing programs. *Z. Kristallogr.* **221**, 15–27 (2006).
- Aroyo, M. I., Kirov, A., Capillas, C., Perez-Mato, J. M. & Wondratschek, H. Bilbao Crystallographic Server. II. Representations of crystallographic point groups and space groups. *Acta Crystallogr.* **A62**, 115–128 (2006).
- Cao, W. Phenomenological theories of ferroelectric phase transitions. *Br. Ceram. Trans.* **103**, 71–75 (2004).
- Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
- Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
- Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
- Stroppa, A. & Kresse, G. The shortcomings of semi-local and hybrid functionals: what we can learn from surface science studies. *New J. Phys.* **10**, 063020 (2008).

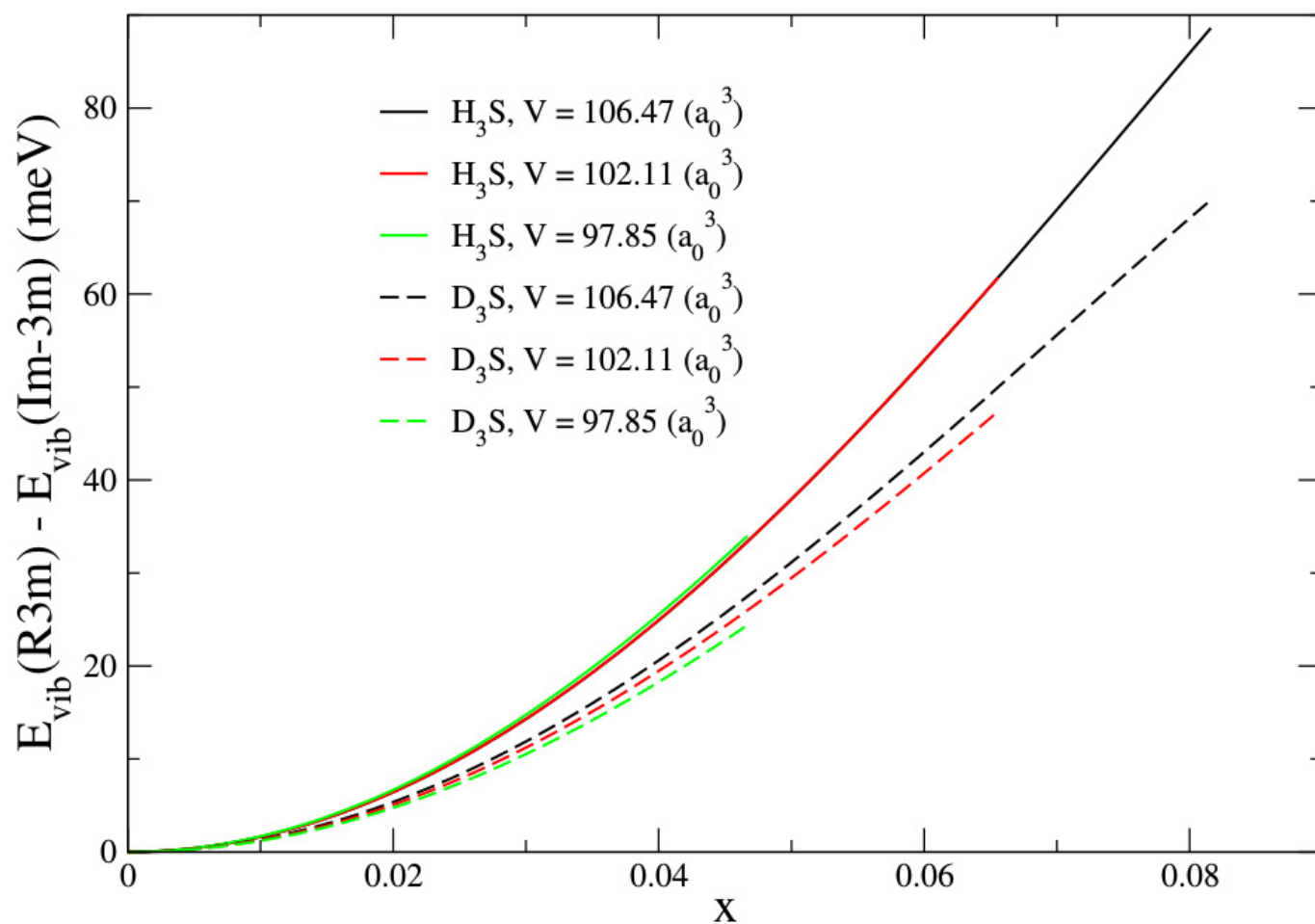


Extended Data Figure 1 | Equations of state. Pressure P as a function of volume V for the *Im-3m* (left) and *R3m* (right) phases, calculated from the static energy E_{BO} that represents the classical nuclei limit (black), and including the vibrational contribution within the SSCHA for both H_3S (red) and D_3S (green).

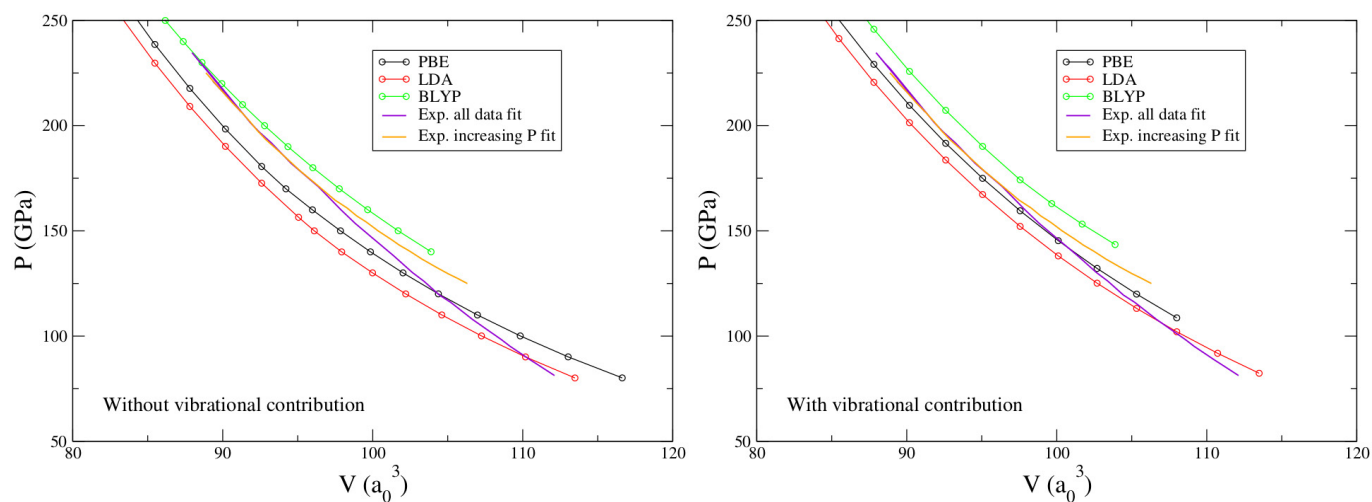


Extended Data Figure 2 | Bond symmetrization within LDA (left) and BLYP (right). For each volume V the relative coordinate $x = (d_2 - a/2)/(a/2)$, in which d_2 is the length of the hydrogen bond and a is the lattice parameter, obtained at the energy minimum is given. When $x = 0$, the covalent and hydrogen bonds have the same length and the structure is

fully symmetric. The energy is calculated at the static level without any vibrational contribution as derived from the BOES (black), and including the quantum anharmonic vibrational contribution both for H_2S (green) and D_3S (purple). The pressure P below which the cubic structure distorts is given in each case.

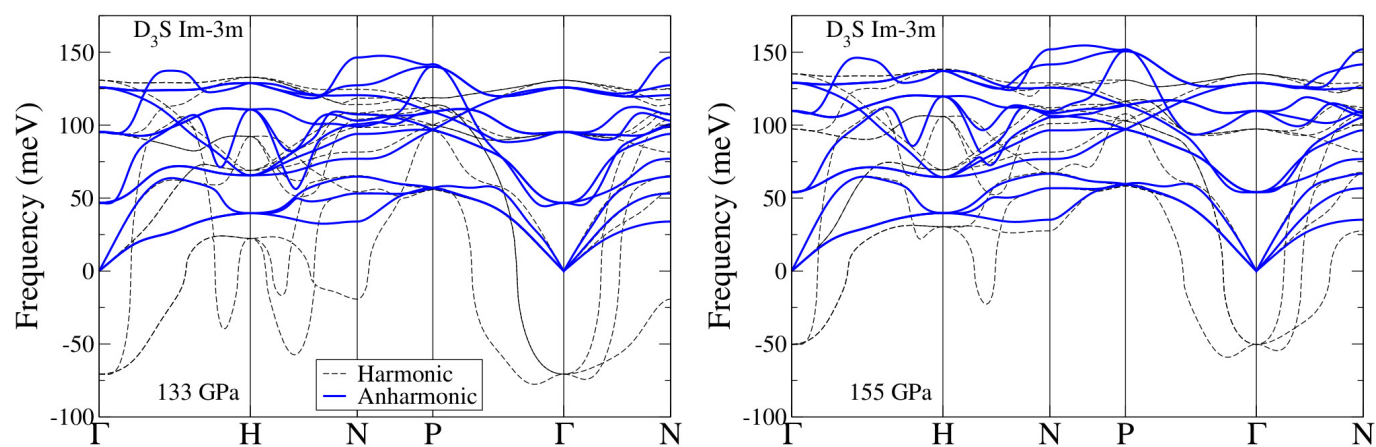


Extended Data Figure 3 | Vibrational energy. SSCHA anharmonic vibrational energy ($E_{\text{vib}}^{\text{R3m}}(x) - E_{\text{vib}}^{\text{Im3m}}$) calculated as a function of the relative coordinate $x = (d_2 - a/2)/(a/2)$ for different volumes (see legend).

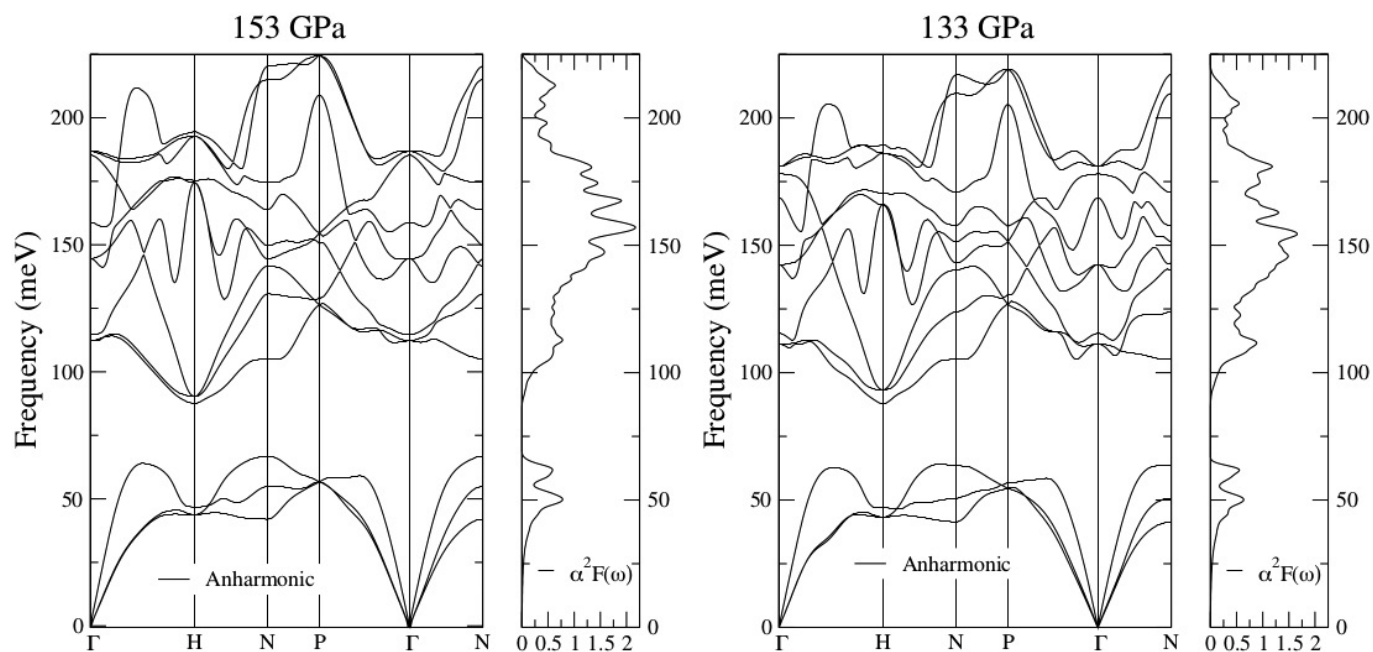


Extended Data Figure 4 | Dependence of the equation of state on the density functional. The equation of state is calculated with different exchange correlation functionals with (right) and without (left) the vibrational contribution to the pressure. At each volume V the pressure P

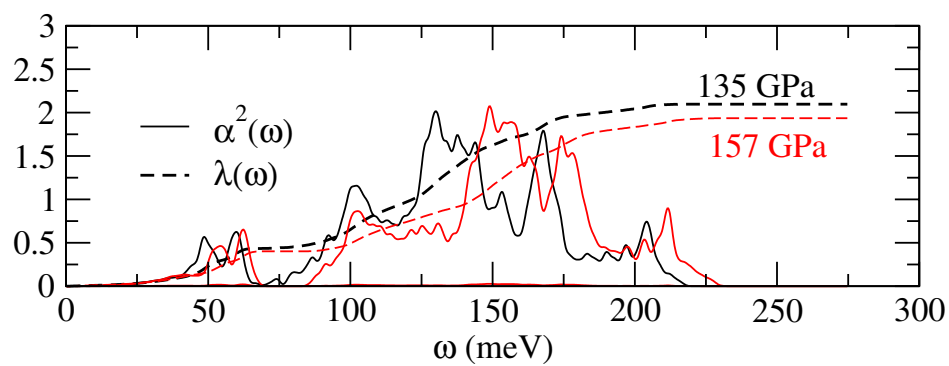
of the structure with minimum energy is given, which depends on whether or not the vibrational contribution is included (see Extended Data Fig. 2). The results are compared with the two curves obtained experimentally¹⁰.



Extended Data Figure 5 | Phonons of $Im\bar{3}m$ D_3S . Comparison between the harmonic and anharmonic phonons of $Im\bar{3}m$ at two different pressures for D_3S : 133 GPa (left) and 155 GPa (right).



Extended Data Figure 6 | Anharmonic phonons of $R3m$ H_3S . Anharmonic phonons are shown at two different pressures: 153 GPa (left) and 133 GPa (right). The Eliashberg functions $\alpha^2 F(\omega)$ are also shown.



Extended Data Figure 7 | Superconducting properties of $Im\bar{3}m$ H_3S . Anharmonic Eliashberg function $\alpha^2F(\omega)$ (solid lines) and integrated electron–phonon coupling constant $\lambda(\omega)$ (dashed lines) of the $Im\bar{3}m$ phase as functions of frequency ω at two different pressures: 135 GPa (black) and 157 GPa (red).

Extended Data Table 1 | Birch–Murnaghan fit to the equation of state

	B_0 (GPa)	V_0 (a_0^3)	B_0'
Classical nuclei	129.8	158.4	3.6
H ₃ S	86.63	176.4	3.9
D ₃ S	87.81	174.8	3.9

Parameters of the fit to $P(V) = \frac{3}{2}B_0\left[(V_0/V)^{7/3} - (V_0/V)^{5/3}\right]\left\{1 + \frac{3}{4}(B_0' - 4)\left[(V_0/V)^{2/3} - 1\right]\right\}$ for the equation of state of the $Im\bar{3}m$ phase with classical nuclei, and with vibrational contributions for both H₃S and D₃S. B_0 , bulk modulus at zero pressure; V_0 , reference volume; B_0' , pressure derivative of the bulk modulus at zero pressure.

Extended Data Table 2 | Raman and infrared active modes

<i>Im3m</i>							
Mode	Degeneracy	R or I active?	ω (meV)				
			H ₃ S		D ₃ S		
			135 GPa	157 GPa	133 GPa	155 GPa	
T_{1u}	3	I	79.6	87.6	46.7	54.1	
T_{1u}	3	I	131.0	152.1	95.4	109.8	
T_{2u}	3		177.3	182.0	125.8	129.0	
<i>R3m</i>							
Mode	Degeneracy	R or I active?	ω (meV)				
			H ₃ S		D ₃ S		
			133 GPa	153 GPa	133 GPa	153 GPa	
E	2	R+I	111.3	112.2	74.4	73.7	
A_1	1	R+I	115.4	114.7	79.8	73.7	
E	2	R+I	142.4	144.6	105.6	105.9	
A_1	1	R+I	168.7	158.8	123.2	115.0	
A_2	1		178.0	185.3	125.1	130.7	
E	2	R+I	181.0	186.7	127.0	131.7	

Phonon modes at the Γ point for the *Im $\bar{3}m$* and *R3m* phases at different pressures calculated including anharmonicity within the SSCHA. The results for *R3m* are obtained keeping the centroid position at the minimum of the BOES. Raman (R) and infrared (I) activity is indicated for each mode.

Extended Data Table 3 | Superconducting parameters

	Compound	P (GPa)	λ	ω_{log} (meV)	T_c (K)
SSCHA $Im\bar{3}m$	H ₃ S	135	2.10	104.9	214
	H ₃ S	157	1.94	113.8	216
	D ₃ S	133	2.45	70.5	163
	D ₃ S	155	2.08	82.9	163
SSCHA $R\bar{3}m$	H ₃ S	133	1.62	105.8	163
	H ₃ S	153	1.93	109.3	205
	D ₃ S	132	1.64	79.8	123
	D ₃ S	152	2.00	81.4	155

Calculated λ , ω_{log} and T_c . The pressure given includes the vibrational contribution.

Continuous directional water transport on the peristome surface of *Nepenthes alata*

Huawei Chen^{1*}, Pengfei Zhang^{1*}, Liwen Zhang¹, Hongliang Liu², Ying Jiang³, Deyuan Zhang¹, Zhiwu Han⁴ & Lei Jiang^{2,3}

Numerous natural systems contain surfaces or threads that enable directional water transport^{1–7}. This behaviour is usually ascribed to hierarchical structural features at the microscale and nanoscale, with gradients in surface energy^{8,9} and gradients in Laplace pressure¹⁰ thought to be the main driving forces. Here we study the prey-trapping pitcher organs of the carnivorous plant *Nepenthes alata*. We find that continuous, directional water transport occurs on the surface of the ‘peristome’—the rim of the pitcher—because of its multiscale structure, which optimizes and enhances capillary rise^{11,12} in the transport direction, and prevents backflow by pinning in place any water front that is moving in the reverse direction. This results not only in unidirectional flow despite the

absence of any surface-energy gradient, but also in a transport speed that is much higher than previously thought. We anticipate that the basic ‘design’ principles underlying this behaviour could be used to develop artificial fluid-transport systems with practical applications.

The tropical plant *N. alata* has a ‘pitcher’—a highly modified, prey-trapping organ, filled with digestive fluid—that is 10–15 centimetres long; at the upper rim is an arch-shaped ring of tissue known as the peristome, which is about 1–2 centimetres wide from the inner to the outer side (Fig. 1a). The pitcher can digest insects to meet fundamental nutrient needs^{13,14}, after capturing these insects when they ‘aquaplane’¹⁵ on the peristome surface, which is completely wettable

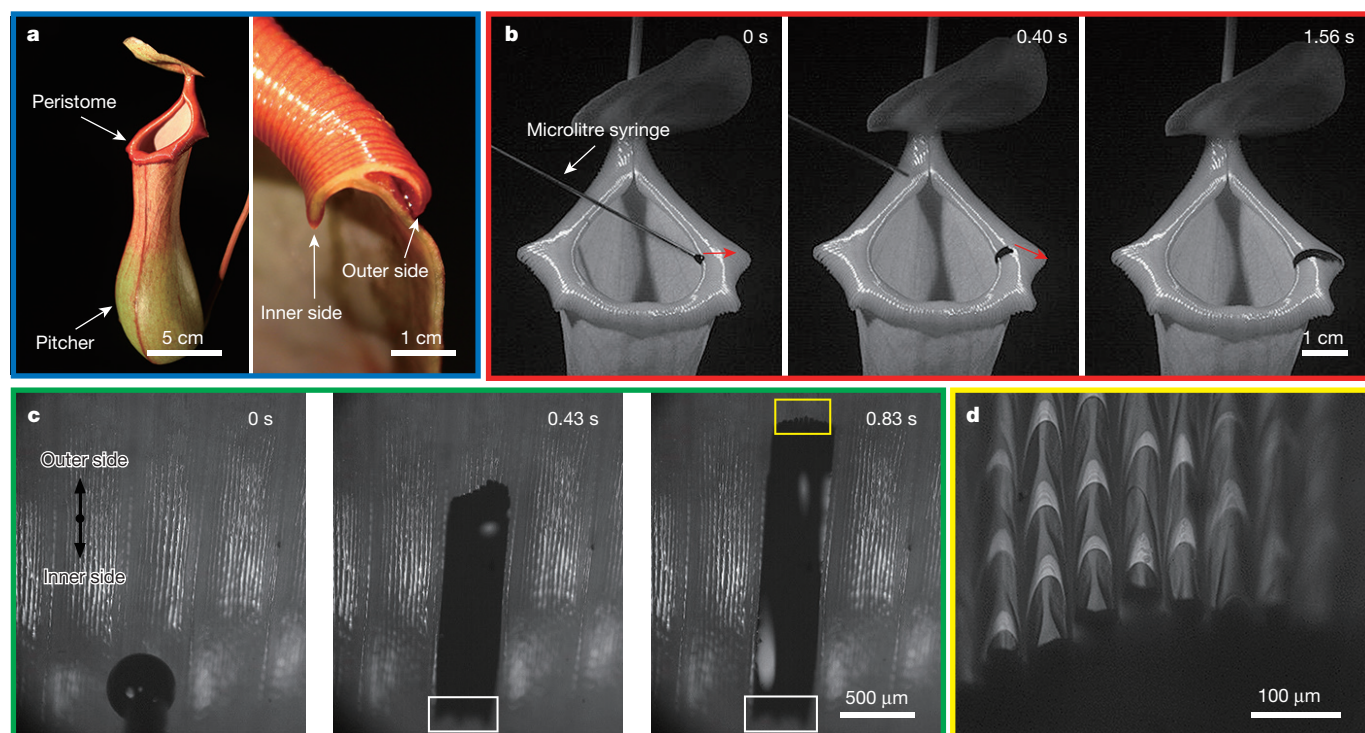


Figure 1 | *In situ* optical observation of directional water transport on the peristome surface. **a**, Optical images of a pitcher of *N. alata*, showing a conspicuous prey-trapping peristome (left) and a cross-sectional image of the peristome (right). The pitcher is 10–15 cm long, with the peristome at its upper rim. This peristome is arch-shaped, and the distance from the inner to the outer side is approximately 1–2 cm. **b**, A water droplet (mixed with blue ink) is dropped onto the peristome with a microlitre syringe; this droplet can overcome gravity to be directionally transported from the inner side to the outer side (but not in the opposite direction) within

several seconds. The red arrows show the direction of water transport. **c**, Water transport from the inner side to the outer side is confined within a single large channel. The starting boundary line (marked by white rectangles) of deposited water does not change during transport. **d**, This enlarged partial image of the water boundary indicates that water transport is further confined within approximately ten microchannels, distributed within the large channel. Each microchannel consists of arch-shaped microcavities with a spacing of approximately 100 μm. The black parts in **b–d** mark the water droplet.

¹School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China. ²Laboratory of Bio-inspired Smart Interface Science, Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100190, China. ³School of Chemistry and Environment, Beihang University, Beijing 100191, China. ⁴Key Laboratory for Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China.

*These authors contributed equally to this work.

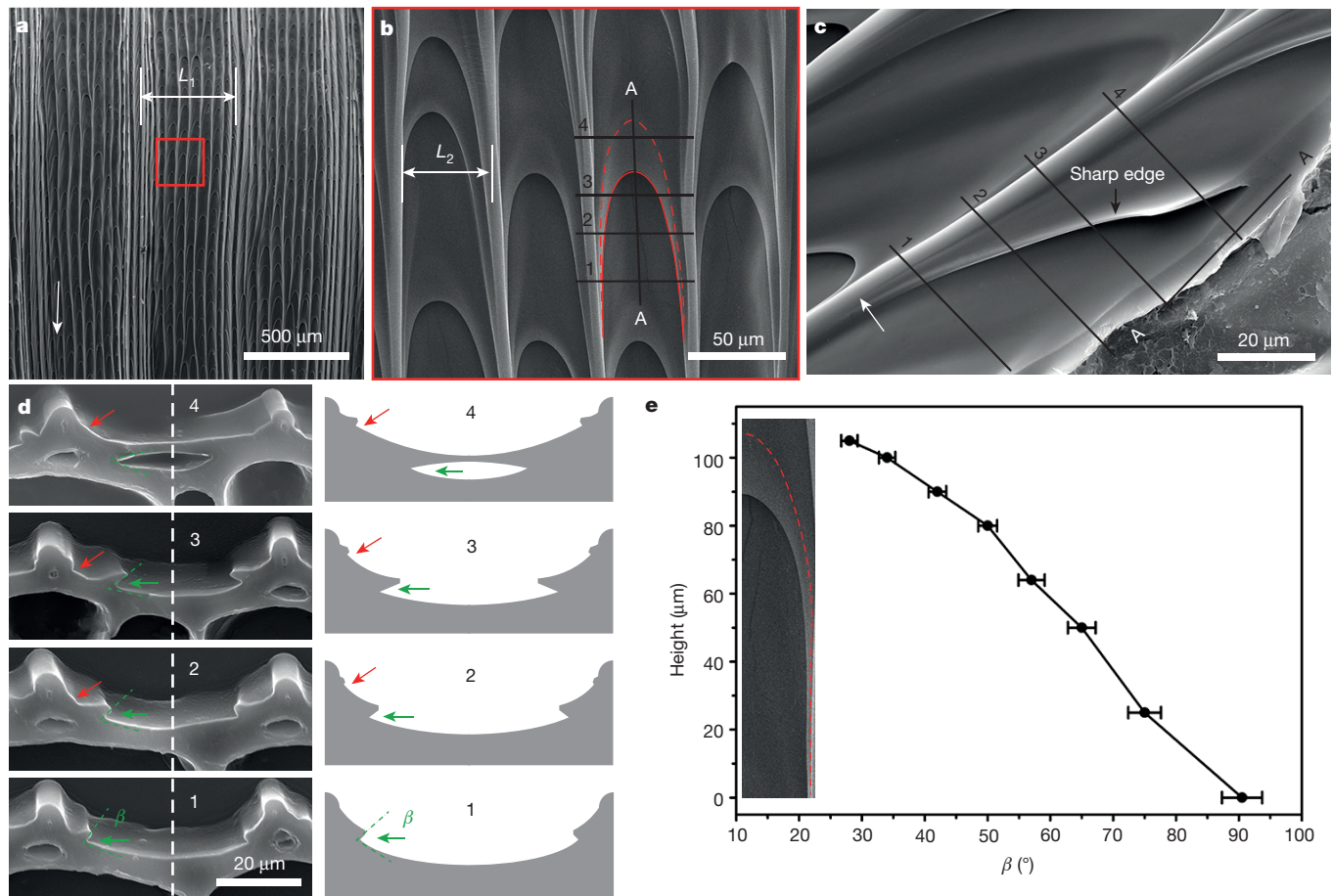


Figure 2 | Surface features of the peristome in the pitchers of *N. alata*. **a**, The peristome surface exhibits a unique structural feature of two-order microgrooves. The width, L_1 , of the first-order microgroove is $461.72 \pm 49.93 \mu\text{m}$. The white arrow indicates the direction of the inner side of the peristome. **b**, Each first-order microgroove consists of approximately ten second-order microgrooves, each containing periodic duck-billed microcavities with arch-shaped edges. The width, L_2 , of the second-order microgroove at the valley of the first-order microgroove is $50.18 \pm 6.18 \mu\text{m}$. The numbers 1–4 correspond to the same numbers in panel **d**. **c**, A vertical section taken at the A–A position in panel **b** reveals that the duck-billed microcavity is closed at the top, with a sharp edge,

by water from rain, dew or high humidity and by nectar secreted from extrafloral nectaries at the pitcher's inner margin^{16,17}. Slippery surfaces have been designed to mimic the structure and behaviour of the natural peristome surface^{18,19}, but questions remain about the mechanism(s) underlying its function.

Here, we use high-speed camera observations to monitor water movement along the peristome surface (Fig. 1b). We find directional transport from the inner to the outer side: the water of a droplet deposited on the inner peristome margin moves to the outer margin within several seconds, while a water droplet placed on the outer margin cannot move towards the inside. *In situ* optical microscopic observations (Fig. 1c) show that water transport is confined within a single large channel running perpendicular to the peristome rim; the water does not spread laterally beyond the width of the initial deposition area (marked by white rectangles in Fig. 1c; see also Supplementary Video 1; for water transport along several parallel large channels, see Supplementary Video 2). The higher-resolution image in Fig. 1d shows that water transport is in fact confined to approximately ten microchannels distributed within the large channel, and that each microchannel consists of arch-shaped microcavities with a spacing of roughly $100 \mu\text{m}$. These structural features ensure that water evaporating from the pitcher and condensing on the inner margin will spread over the

entire peristome surface to make it more slippery, and are also likely to cause the directionality of the water transport.

Detailed structural characterization of the peristome surface with a scanning electron microscope (SEM) reveals a two-order hierarchical arrangement of regular radial ridges, aligned so that they form a two-order hierarchy of parallel microgrooves (Fig. 2a; see Extended Data Fig. 1a for a cross-section). Each first-order microgroove—with a width, L_1 , of $461.72 \pm 49.93 \mu\text{m}$ (Fig. 2a)—contains about ten second-order microgrooves whose width, L_2 , increases from the ridge to the valley of the first-order microgroove (reaching $50.18 \pm 6.18 \mu\text{m}$ at the valley). Arch-shaped (or duck-billed) microcavities are regularly distributed along the second-order microgrooves (Figs 1d and 2b), with an overall upward slant and with the tops of the arches pointing towards the outer side. A vertical section through the peristome shows that the top sections of the microcavities are closed (Fig. 2c; see Extended Data Fig. 1b for side view), and that the enclosing surface is slightly slanted and has a sharp edge. The periodic microcavities overlap along the ridge (the starting point of which is marked by a white arrow in Fig. 2c), with the overlap between a given microcavity and its overlying neighbour becoming more evident on moving towards the top of the microcavity (red arrows in Fig. 2d). The sides of each cavity form a wedge (green arrows in Fig. 2d), with the opening angle, β , declining

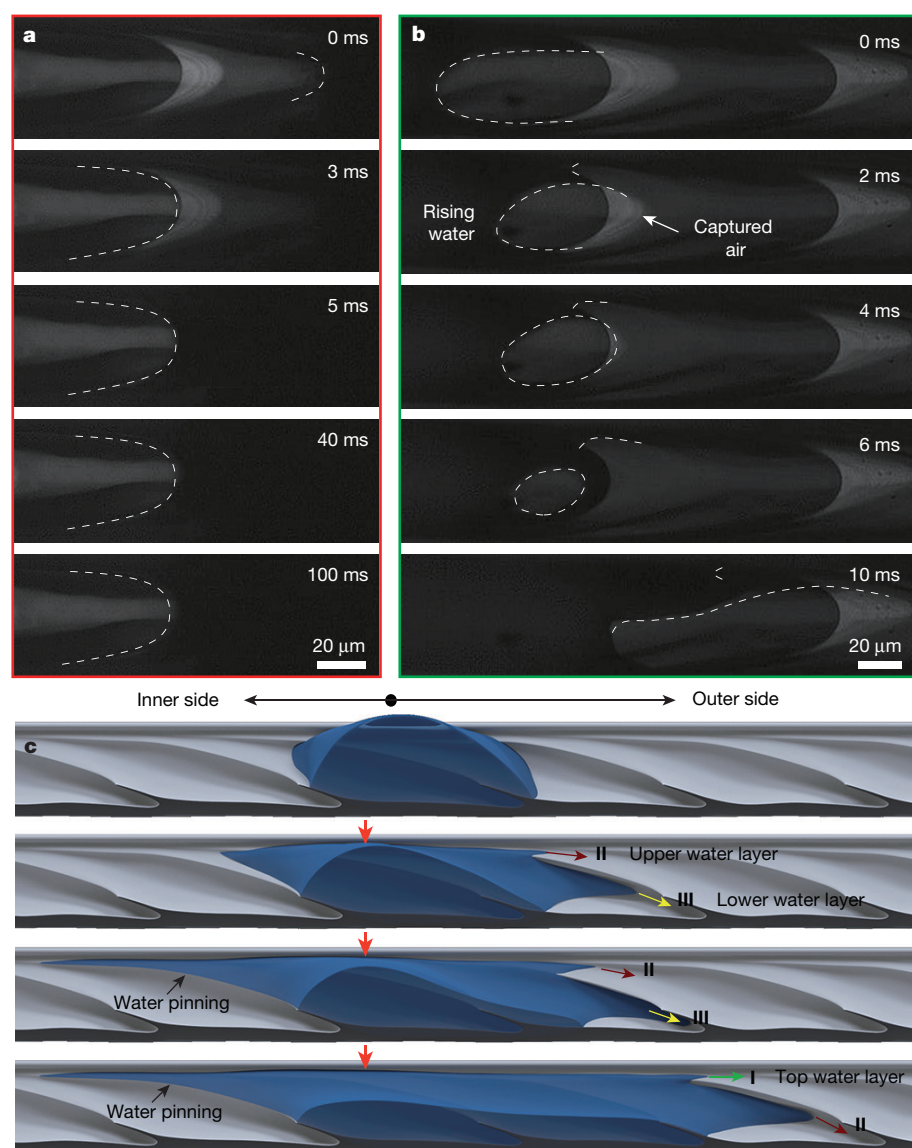


Figure 3 | Directional water-transport processes on the peristome surface. **a**, High-speed digital images showing how the starting boundary of a water droplet (dashed lines) changes after the droplet is deposited on the peristome surface. The final boundary coincides with the outline of the microcavity, indicating that water spreading was halted by the microcavity's sharp edge. Water spreading from the outer to the inner side of the peristome is stopped by being pinned at the sharp edge. **b**, High-speed digital images of the continuous water-transport process between overlapping duck-billed microcavities. As the water fills one microcavity, some water overflows to begin filling the next. The dashed lines show the boundary of water transport. **c**, Three-dimensional illustration of the water-transport process. The lower water layer, layer III, fills up a single microcavity (yellow arrows) and overflows to generate the upper water layer, layer II. Before the second microcavity is completely filled, the upper water layer (brown arrows) becomes the top water layer, layer I (green arrows), filling the third microcavity. Transport in the reverse direction does not occur because the water is pinned in place, with the water boundary coinciding with the sharp edge.

from an initial value of approximately 90° at the bottom of the microcavity to roughly 28° at the top (Fig. 2e and Extended Data Fig. 2).

In situ observation of the water-spreading process along a single second-order microgroove shows that the water contact line (indicated by dashed lines in Fig. 3a) after deposition coincides with the outline of the microcavity, indicating that water is pinned at the sharp edge of the microcavity. The edge angle, ϕ (the angle subtended by the two surfaces that form the sharp edge), of approximately 2° to 8° favours water pinning, according to the Gibbs inequality²⁰, and prevents wetting from proceeding from the outer to the inner side of the peristome (Extended Data Fig. 1b; see Methods for further discussion). In contrast, water transport from the inner to the outer side occurs simply through continuous filling of a single microcavity. The outline of the water boundary (indicated by the dashed lines in Fig. 3b) shows that water initially spreads along the corner of the wedge, then pushes air out to fill the microcavity, and finally converges at the front of the microcavity (Fig. 3b and Supplementary Video 3; see Extended Data Fig. 3 for a complete illustration of water filling a single microcavity). Because neighbouring microcavities overlap, the filling of one microcavity begins before that of the preceding microcavity is completed.

The overall result is continuous and directional water transport, as illustrated in Fig. 3c: when a given microcavity fills to the extent that

its water level exceeds the top of its confining side wall, around the starting point of the overlap with the next microcavity (white arrow in Fig. 2c), the lower water layer (layer III; yellow arrows) overflows and generates the upper water layer (layer II), which then begins to fill the second microcavity (brown arrows). Before the second microcavity is full, the cycle repeats and the upper water layer (layer II) overflows to generate the top water layer (layer I; green arrow), which fills the third microcavity. Sequential filling of each microcavity achieves continuous water transport, while water pinning prevents wetting from occurring in the reverse direction.

Wettability measurements show that the peristome surface is hydrophilic, and energy-dispersive spectrometer (EDS) measurements indicate the lack of a chemical gradient (Extended Data Fig. 4). To explore the role of surface energy further, we made artificial peristomes using a polydimethylsiloxane (PDMS) replica moulding method²¹ (Extended Data Fig. 5) and adjusted their surface energy using oxygen plasma treatment²². We detected directional water transport only with PDMS surfaces that are hydrophilic (Fig. 4a), with an intrinsic contact angle between the water and the surface of less than about 65° , which constitutes the newly proposed boundary^{23,24} between hydrophobicity and hydrophilicity (the previously known contact-angle boundary was 90°). The artificial PDMS peristome with a superhydrophilic surface displayed the fastest water-spreading

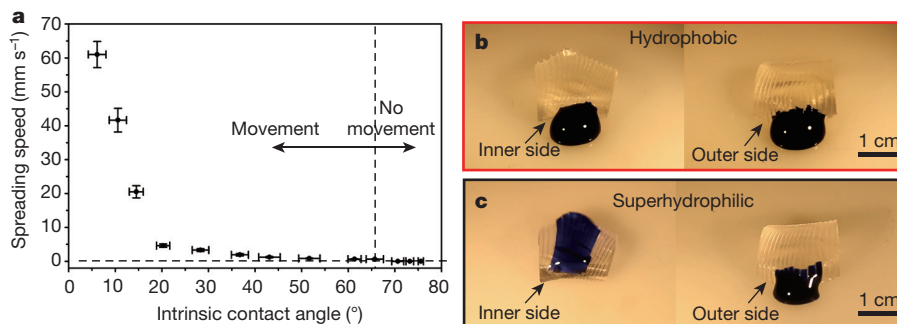


Figure 4 | Mechanism of directional water transport. **a**, Directional water transport occurred on the artificial PDMS replica only when the PDMS was hydrophilic (that is, when the intrinsic contact angle between the water and the PDMS molecules was less than 65°); the water-transport speed was maximized when the surface was modified to be superhydrophilic (such that the intrinsic contact angle was less than 5°). Error bars indicate standard deviations from at least five independent

speed, close to the transport speed measured for the natural peristome surface (approximately $78 \pm 12 \text{ mm s}^{-1}$) (Fig. 4a). Figure 4b, c illustrates the transport experiments. No water transport occurred on the hydrophobic PDMS surface when water was placed on either the inner side or the outer side (Fig. 4b and Extended Data Fig. 6c, d). However, when the artificial surface was superhydrophilic, directional water transport occurred (Fig. 4c and Extended Data Fig. 6e, f) much as it does as on the natural peristome (Extended Data Fig. 6a, b). Superhydrophilicity is thus crucial for directional water transport in this system, and ensures that transport occurs on both the natural and the artificial peristome surface much more quickly than was seen in an earlier study documenting unidirectional wetting caused by an asymmetric surface structure²⁵.

Superhydrophilic surfaces generally have no capacity for directional water movement unless they contain a chemical gradient^{8,9}. In the case of the peristome surface, which lacks such a gradient, the directional water transport arises from its unique structural features, which enable water to fill microcavities because of capillary rise at corners, as first described in 1712 (refs 11, 12). Specifically, this type of capillary rise occurs at the corner formed by two vertical intersecting hydrophilic plates corresponding to the wedge corner of the microcavity shown in Fig. 2c (Extended Data Fig. 7a, right), which the liquid can fill and use to rise unbounded provided that the Young contact angle, θ , of the liquid on the plate and the corner opening angle, α , satisfy $\alpha/2 + \theta \leq 90^\circ$ (refs 26, 27). The capillary rise height, $H_c(x)$, is in this case given by the following equation^{28–30}:

$$H_c(x) = \frac{2\gamma \cos\theta}{\rho g x}$$

where γ , ρ , θ and g are the surface tension, the density of the liquid, the water contact angle and the gravitational constant, respectively. However, nature does not use simple corner capillary rise with a constant opening angle, but renders it more powerful by structuring the peristome so that there is a gradient: the opening angle α_1 at the bottom of a given microcavity is larger than the opening angle α_2 at its top (Extended Data Fig. 2). For such a case of capillary rise, when $\alpha_1 > \alpha_2$ (Extended Data Fig. 7b, right), and assuming that the difference between α_1 and α_2 is infinitesimal (see Methods), the capillary height rise $H_c(x)$ can be presented as:

$$H_c(x) = \frac{2\gamma \cos\theta}{\rho g x \alpha_1} + \frac{\alpha_1 - \alpha_2}{\alpha_1 h} \frac{4\gamma^2 \cos^2\theta}{\rho^2 g^2 x^2 \alpha_1^2} + \dots$$

where h is the height of the two intersecting plates. Comparison of the two equations shows that the capillary rise height in the case of a

constant opening angle is obtained by setting $\alpha_1 = \alpha_2$, and that the capillary rise height is larger (or smaller) in the case of $\alpha_1 > \alpha_2$ (or $\alpha_1 < \alpha_2$). Simple experiments (Extended Data Fig. 7) confirm the enhancement and reduction in water rise caused by the gradient of the opening angle and the opposite gradient. Interestingly, in a system in which the vertical corner leads to a top with a closed horizontal corner, water ascends along the vertical gradient and then continues to fill the top horizontal corner. This effect increases the volume of water that rises in the system by about 40% (Extended Data Fig. 8a), and prolongs the water-retention time to twice that seen in the system with a fixed opening angle (Extended Data Fig. 8b). In the peristome, the use of a microcavity structure that is symmetrical and closed at the top taps into these effects, and enables a much stronger water rise than would be possible by ordinary capillary rise. This, in conjunction with water pinning at the sharp edges of the microcavity in the reverse direction, enables continuous and fast directional transport of water along the peristome surface. Artificial systems exploiting these principles might find use in applications that call for directional fluid transport—for example, in agricultural drip irrigation or in the non-powered delivery of microdrugs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 February 2015; accepted 25 January 2016.

- Parker, A. R. & Lawrence, C. R. Water capture by a desert beetle. *Nature* **414**, 33–34 (2001).
- Zheng, Y. M. *et al.* Directional water collection on wetted spider silk. *Nature* **463**, 640–643 (2010).
- Ju, J. *et al.* A multi-structural and multi-functional integrated fog collection system in cactus. *Nature Commun.* **3**, 1247 (2012).
- Ishii, D. *et al.* Water transport mechanism through open capillaries analyzed by direct surface modifications on biological surfaces. *Sci. Rep.* **3**, 3024 (2013).
- Bai, H. *et al.* Direction controlled driving of tiny water drops on bioinspired artificial spider silks. *Adv. Mater.* **22**, 5521–5525 (2010).
- Ju, J., Zheng, Y. M. & Jiang, L. Bioinspired one-dimensional materials for directional liquid transport. *Acc. Chem. Res.* **47**, 2342–2352 (2014).
- Li, K. *et al.* Structured cone arrays for continuous and effective collection of micron-sized oil droplets from water. *Nature Commun.* **4**, 2276 (2013).
- Chaudhury, M. K. & Whitesides, G. M. How to make water run uphill. *Science* **256**, 1539–1541 (1992).
- Daniel, S., Chaudhury, M. K. & Chen, J. C. Past drop movements resulting from the phase change on a gradient surface. *Science* **291**, 633–636 (2001).
- Lorenceanu, E. & Quéré, D. Drops on a conical wire. *J. Fluid Mech.* **510**, 29–45 (2004).
- Taylor, B. Concerning the ascent of water between two glass plates. *Phil. Trans. R. Soc. Lond.* **27**, 538 (1712).
- Hauksbee, F. An experiment touching the ascent of water between two glass plates in a hyperbolic figure. *Phil. Trans. R. Soc. Lond.* **27**, 539 (1712).
- Juniper, B. E. & Burras, J. K. How pitcher plants trap insects. *New Sci.* **13**, 75–77 (1962).

14. Ellison, A. M. Nutrient limitation and stoichiometry of carnivorous plants. *Plant Biol.* **8**, 740–747 (2006).
15. Bohn, H. F. & Federle, W. Insect aquaplaning: *Nepenthes* pitcher plants capture prey with the peristome, a fully wettable water-lubricated anisotropic surface. *Proc. Natl Acad. Sci. USA* **101**, 14138–14143 (2004).
16. Bauer, U. & Federle, W. The insect-trapping rim of *Nepenthes* pitchers: surface structure and function. *Plant Signal. Behav.* **4**, 1019–1023 (2009).
17. Bauer, U., Bohn, H. F. & Federle, W. Harmless nectar source or deadly trap: *Nepenthes* pitchers are activated by rain, condensation and nectar. *Proc. R. Soc. B* **275**, 259–265 (2008).
18. Wong, T. S. *et al.* Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity. *Nature* **477**, 443–447 (2011).
19. Lafuma, A. & Quéré, D. Slippery pre-suffused surfaces. *EPL* **96**, 56001 (2011).
20. Oliver, J. F., Huh, C. & Mason, S. G. Resist to spreading of liquids by sharp edges. *J. Colloid Interface Sci.* **59**, 568–581 (1977).
21. Sun, M. H. *et al.* Artificial lotus leaf by nanocasting. *Langmuir* **21**, 8978–8981 (2005).
22. Zhou, J. W., Ellis, A. V. & Voelcker, N. H. Recent developments in PDMS surface modification for microfluidic devices. *Electrophoresis* **31**, 2–16 (2010).
23. Vogler, E. A. Structure and reactivity of water at biomaterial surfaces. *Adv. Colloid Interface Sci.* **74**, 69–117 (1998).
24. Tian, Y. & Jiang, L. Wetting: intrinsically robust hydrophobicity. *Nature Mater.* **12**, 291–292 (2013).
25. Chu, K.-H., Xiao, R. & Wang, E. N. Uni-directional liquid spreading on asymmetric nanostructured surfaces. *Nature Mater.* **9**, 413–417 (2010).
26. Concus, P. & Finn, R. On the behavior of a capillary surface in a wedge. *Proc. Natl Acad. Sci. USA* **63**, 292 (1969).
27. Finn, R. Capillary surface interfaces. *Not. Am. Math. Soc.* **46**, 770–781 (1999).
28. Lopez de Ramos, A. & Cerro, R. L. Liquid filament rise in corners of square capillaries: a novel method for the measurement of small contact angles. *Chem. Eng. Sci.* **49**, 2395–2398 (1994).
29. Higuera, F. J., Medina, A. & Linan, A. Capillary rise of a liquid between two vertical plates making a small angle. *Phys. Fluids* **20**, 102102 (2008).
30. Ponomarenko, A., Quéré, D. & Clanet, C. A universal law for capillary rise in corners. *J. Fluid Mech.* **666**, 146–154 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the National Natural Science Foundation of China (grant nos 51290292, 51175020, 51475029 and 21431009) and the Innovation Foundation of the Beijing University of Aeronautics and Astronautics (BUAA; now Beihang University) for PhD Graduates for providing financial support. We also thank G. Wang and Y. Lai from the National Natural Science Foundation of China, and S. Wang from Tianjin University, for support and discussions.

Author Contributions H.C., P.Z. and L.Z. performed the experiments: H.C. and P.Z. worked on the water transport and the characterization of *Nepenthes alata*; H.C., P.Z. and L.Z. investigated the capillary rise in the wedge; P.Z. fabricated the artificial *Nepenthes* surface. H.C., P.Z., L.Z., D.Z. and L.J. collected and analysed the data and proposed the mechanism of water transport on the peristome surface. Y.J. carried out the theoretical analysis of capillary rise in the gradient wedge. H.C., P.Z., Z.H., H.L., D.Z. and L.J. wrote the text. H.C., D.Z. and L.J. conceived the project and designed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D. Zhang (zhangdy@buaa.edu.cn), H. Chen (chenhw75@buaa.edu.cn) or L. Jiang (jianglei@iccas.ac.cn).

METHODS

Materials. Mature pitcher plants, *N. alata*, growing in a greenhouse were purchased commercially from Kunjiyuanyi Corporation (Guanddong province, China) and carefully maintained, covered with plastic film, in a small greenhouse under controlled conditions at a temperature of 20 °C to 25 °C and humidity of 75% to 90%.

Characterization of the peristome surface. We recorded optical images of the pitcher of *N. alata* using a digital camera (600D, Canon, Japan). We observed the microstructures of the peristome surface using a SEM (CamScan-3400, CamScan Corp., UK) operating at a 20 kV acceleration voltage. To avoid severe distortion caused by dehydration, we used a freeze-drying method to prepare the SEM specimens. Wettability was investigated by depositing a water droplet (approximately 4 µl) on the peristome surface, and the water-spreading process was monitored using an optical angle measuring system (SL200B, Solon, China) (Extended Data Fig. 4a). The chemical composition of the peristome surface was determined by EDS equipped with the SEM (Extended Data Fig. 4b).

Observation of water transport. We used a high-speed camera (I-speed LT, Olympus, Japan) to record the water-transport process. For our initial observations of the macroscopic water-transport process (Fig. 1b), we first deposited a water droplet (mixed with blue ink, roughly 5 µl in total) on the inner margin of the peristome surface using a microlitre syringe; then, we cut a partial peristome out of a pitcher and placed it under the eyepiece of an optical microscope (BX51, Olympus, Japan). For the observations reported in Fig. 1c, d and Supplementary Video 1, we used a microsyringe to deposit a small amount of water (about 0.1 µl) on the inner surface of the peristome surface, and investigated the water-spreading process both in the microgrooves and in the duck-billed microcavity. When observing water spreading in two separate large channels (see Supplementary Video 2), we deposited about 0.2 µl of water.

To observe the water-spreading process along a single second-order microgroove (see Supplementary Video 3, Fig. 3b and Extended Data Fig. 3), we deposited approximately 0.01 µl of water on the inner margin of the peristome and monitored the top boundary of the water. To observe water pinning at the sharp edge of the microcavity, we deposited roughly 0.01 µl of water on the peristome surface and monitored the starting boundary of the water with the high-speed camera (Fig. 3a). The water contact line was identical to the sharp edge of the microcavity, indicating that water was pinned by the sharp edge running in the direction from the outer to the inner side of the peristome. To investigate directional water transport on the peristome and its artificial surfaces, we placed the samples on a polyethylene plate and then deposited water on one side. Sequential images of water transport are shown in Extended Data Fig. 6.

Fabrication of, and water spreading on, artificial replicas. We used a simple replica moulding method to fabricate artificial peristomes on an intact natural pitcher using polydimethylsiloxane (PDMS). SEM images (Extended Data Fig. 5) showed that all of the surface structural features of the peristome were accurately transferred to the artificial replica. The prepolymer was formed by mixing PDMS and curing agent at a mass ratio of 10/1. Before the prepolymer was cast onto the surface of fresh slippery zone, it was well stirred for 15 min, and then degassed in a vacuum oven for 15 min. The negative replicas of the pitcher were fabricated by conducting the following procedures: casting, curing in a vacuum oven, demoulding. Then the negative replicas were treated by CF₄ plasma to form an anti-sticker layer. A second replication of PDMS was performed on the treated negative replicas in the same way as before, to obtain the replicas of the peristome. (For further details of sample preparation and fabrication, see ref. 21.)

Modification of the PDMS surface by oxygen was conducted using RF plasma (P8C, Schwarze, China) at an RF power of 100 W for 5 min, a system pressure of 100 µbar and a flow rate of 20 s.c.c.m. After oxygen treatment, the PDMS was superhydrophilic, and the contact angle gradually increased with time in air. Both the flat PDMS and the PDMS peristome replicas were fabricated and treated under the same conditions. The change in the contact angle of flat PDMS with time was measured using an optical angle-measuring system (SL200B, Solon, China). Meanwhile, the water-spreading speed on the PDMS replicas (approximately 5 µl water was deposited on the inner side of the replicas) was monitored using the high-speed camera. Thus, we could obtain the relationship between the water-spreading speed on the PDMS replicas and the intrinsic contact angle of PDMS (Fig. 4a).

Water rise along four types of wedge corners. Polyvinylchloride (PVC) plates were used in parallel experiments and were modified to be superhydrophilic by oxygen plasma treatment with an RF power of 100 W for 5 min, a system pressure of 100 µbar and a flow rate of 20 s.c.c.m. The heights of the plates were limited to 100 mm, owing to restricted sizes of the available material. Four types of wedge corners were built (Extended Data Fig. 7): (i) a fixed wedge with a bottom opening angle α_1 equivalent to the top opening angle α_2 ; (ii) a gradient wedge where $\alpha_1 > \alpha_2$ and α_2 is constant at 5°; (iii) a wedge of opposite gradient with $\alpha_1 < \alpha_2$ and α_2 is constant at 90°; and (iv) a top-closed wedge with a blind top, that is, $\alpha_2 = 0^\circ$. The wedge corners were vertically immersed in water mixed with blue ink.

Water-rise volume and water-retention time. To measure the water volume in the four types of wedge corner, we placed the PVC plates vertically in a small dish that perfectly fitted their size. We injected water into the dish to a depth of approximately 3 mm, and placed a round flat floater on the water. When water rose at the wedge corners, the height of the water surface changed. We used a laser displacement sensor with an accuracy of 1 µm to measure the change in height of the water surface, by detecting the distance between the floater and the laser displacement sensor. The relationship between the change in height and the water-rise volume was calibrated by injecting various volumes of water into the dish and measuring the change in height of the floater. The accuracy of the water rise in this experiment was about 8 µl µm⁻¹, and each case was measured five times (Extended Data Fig. 8a).

To measure the time of water volatilization in the four types of wedge corners, we placed the PVC plates vertically in a ventilated heating furnace, and installed a digital camera to observe the amount of water. All of the wedge corners were injected with 5 ml of deionized water at the bottom of the corner. The temperature of the heating furnace was set to 45 °C. We defined the volatilization time as the water-retention time of the wedge corner (Extended Data Fig. 8b).

Mechanism of pinning. Water can be transported from the inner to the outer side of the peristome, but not from the outer to the inner side. This directional pinning probably occurs because of the cooperative effects of the sharp edge (Extended Data Fig. 1b) and the opposite-gradient wedge corners of the microcavity. More specifically, the upper cover of the microcavity has a sharp edge with an angle, ϕ , of approximately 2° to 8°; and the peristome surface is superhydrophilic, with an equilibrium contact angle of about 0°. When the liquid wetting the edge satisfies the condition of Gibbs inequality²⁰:

$$\theta_0 \leq \theta \leq (180^\circ - \phi) + \theta_0$$

(where θ is the contact angle measured through the drop at the edge), the edge will pin the liquid in place. The critical angle of θ at the moment that the liquid contact line just crosses the edge can be calculated to be roughly 172° to 178°. We can see that when θ is smaller than this, the liquid will be pinned by the edge of the microcavity. The value of θ increases as the volume of liquid increases. However, the liquid on the peristome surface can be quickly transported away from the deposited position, and thus it is difficult for θ to reach 172° to 178°; hence, the liquid can be pinned by the edge of the microcavity. When the liquid contact line crosses the edge, the liquid will fill the microcavity. Meanwhile, the opposite-gradient wedge corner can reduce the liquid rise from the outer side to the inner side.

Theoretical analysis of capillary rise in a gradient wedge. When a wedge-shaped gap between intersecting plates is brought into contact with liquid, the liquid can rise along the intersecting corner (Extended Data Fig. 9a). In our system, the height of the liquid surface is determined only by the balance between the interfacial potential ($G_{\text{interface}}$) and the gravitational potential (G_{gravity}). Then the total energy (G_{total}) of the whole system studied here is:

$$G_{\text{total}} = G_{\text{interface}} + G_{\text{gravity}}$$

We choose the section of the two intersecting plates with an infinitely small thickness Δx to analyse the energy potential (Extended Data Fig. 9b).

Interfacial energy

$$G_{\text{interface}} = \left[\frac{H(x)}{\sin \beta_1} + \frac{H(x)}{\sin \beta_2} \right] \times (-\gamma \cos \theta)$$

β is the gradient opening angle. Assuming $\beta_1 \approx \beta_2 \approx 90^\circ$ for simplicity, we get:

$$G_{\text{interface}} = 2H(x) \times (-\gamma \cos \theta)$$

where γ , θ and $H(x)$ represent the surface tension, the contact angle, and the height of the liquid surface at position x , respectively.

Gravitational potential

If:

$$\begin{aligned} G_{\text{gravity}} &= \int_0^{H(x)} \rho \frac{(w_{\text{up}} + w_{\text{down}})}{2} g z dz \\ &= \rho g x \left[\frac{\alpha_1}{2} H^2(x) - \frac{(\alpha_1 - \alpha_2)}{3h} H^3(x) \right] \end{aligned} \quad (1)$$

(where ρ , g and h denote the liquid density, the gravitational acceleration, and the height of the intersecting plates, respectively), then the total energy (G_{total}) is:

$$G_{\text{total}} = 2H(x) \times (-\gamma \cos \theta) + \rho g x \left[\frac{\alpha_1}{2} H^2(x) - \frac{(\alpha_1 - \alpha_2)}{3h} H^3(x) \right]$$

The height of the surface at equilibrium can be readily determined by minimizing G_{tot} with respect to $H(x)$. It leads to:

$$H_e(x) = \frac{2\gamma\cos\theta}{\rho g x \alpha_1} + \frac{\alpha_1 - \alpha_2}{\alpha_1 h} H_e^2(x) \quad (2)$$

And then, we can theoretically compare the liquid rise height in a fixed wedge with that in a gradient wedge, as follows.

For liquid rise in a fixed wedge (that is, when $\alpha_1 = \alpha_2 = \alpha$), we can obtain the liquid rise height:

$$H_e(x) = \frac{2\gamma\cos\theta}{\rho g x \alpha} \quad (3)$$

For liquid in a gradient wedge (when $\alpha_1 > \alpha_2$), then, when the difference between α_1 and α_2 approaches zero (so $\alpha_1 - \alpha_2 \rightarrow 0$), we can approximately obtain the $H_e(x)$ by adopting the first-order correction to equation (2):

$$H_e(x) = \frac{2\gamma\cos\theta}{\rho g x \alpha_1} + \frac{\alpha_1 - \alpha_2}{\alpha_1 h} \frac{4\gamma^2\cos^2\theta}{\rho^2 g^2 x^2 \alpha_1^2} + \dots \quad (4)$$

Because $\alpha_1 > \alpha_2$, the second term at the right hand of equation (4) must be larger than zero. So, comparing equations (3) and (4), we can arrive at the relationship:

$$H_e(x)|_{\alpha_1 > \alpha_2} > H_e(x)|_{\alpha_1 = \alpha_2}$$

For liquid rise in an opposite wedge, where $\alpha_1 < \alpha_2$, we can directly get the relationship:

$$H_e(x)|_{\alpha_1 < \alpha_2} < H_e(x)|_{\alpha_1 = \alpha_2}$$

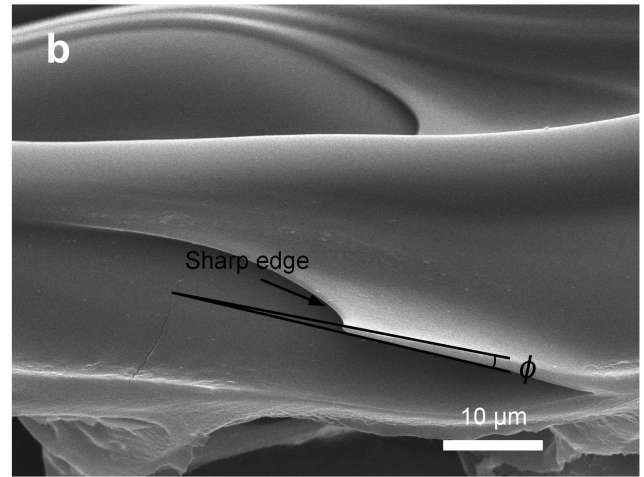
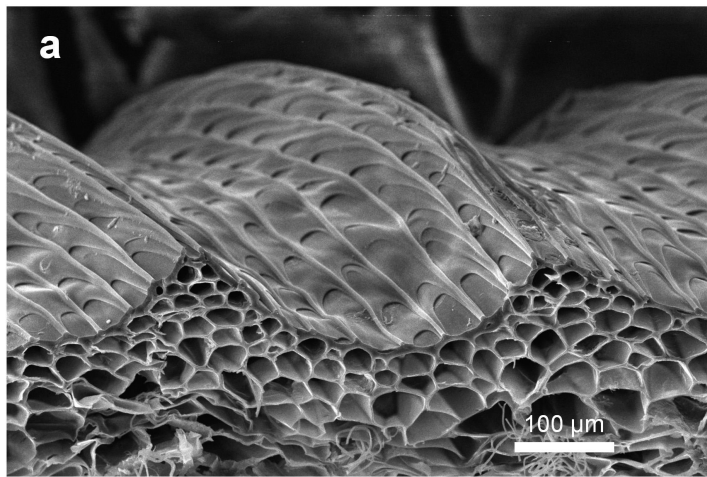
On the basis of this theoretical analysis, we demonstrate that the liquid rise height at the gradient wedge corner (where $\alpha_1 > \alpha_2$) is larger than that at the fixed wedge corner (where $\alpha_1 = \alpha_2$). The results indicate the gradient wedge corner in the microcavity helps the cavity to fill with water and helps with the continuing water transport.

In fact, we can also quantitatively analyse the relative height of the liquid surface for $\alpha_1 \neq \alpha_2$ compared with the case in which $\alpha_1 = \alpha_2$ in a simple way, as follows. In essence, the two components of the total energy of this system play competitive roles in determining the height of the surface: $G_{\text{interface}}$ will raise $H_e(x)$ while G_{gravity} will lower it. If we look at the following term of equation (1), that is, at $-\frac{(\alpha_1 - \alpha_2)}{3h} H^3(x)$, then we can see that, if $\alpha_1 > \alpha_2$, this term will reduce the G_{gravity} , thus:

$$H_e(x)|_{\alpha_1 > \alpha_2} > H_e(x)|_{\alpha_1 = \alpha_2}$$

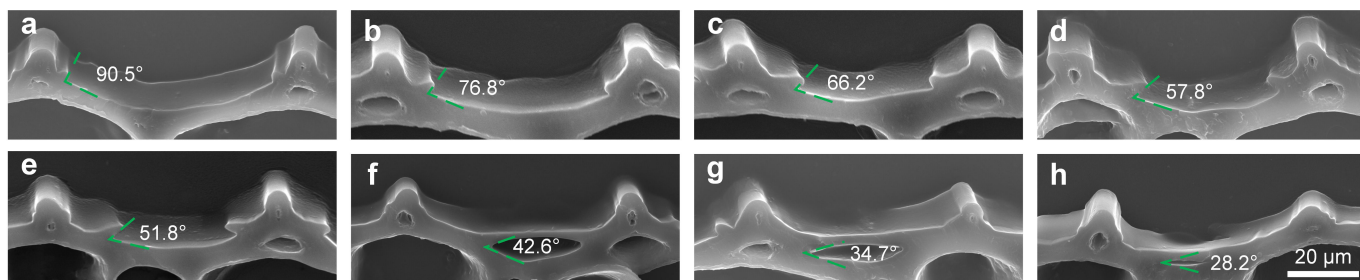
But if $\alpha_1 < \alpha_2$, this term will increase the G_{gravity} , hence:

$$H_e(x)|_{\alpha_1 < \alpha_2} < H_e(x)|_{\alpha_1 = \alpha_2}$$

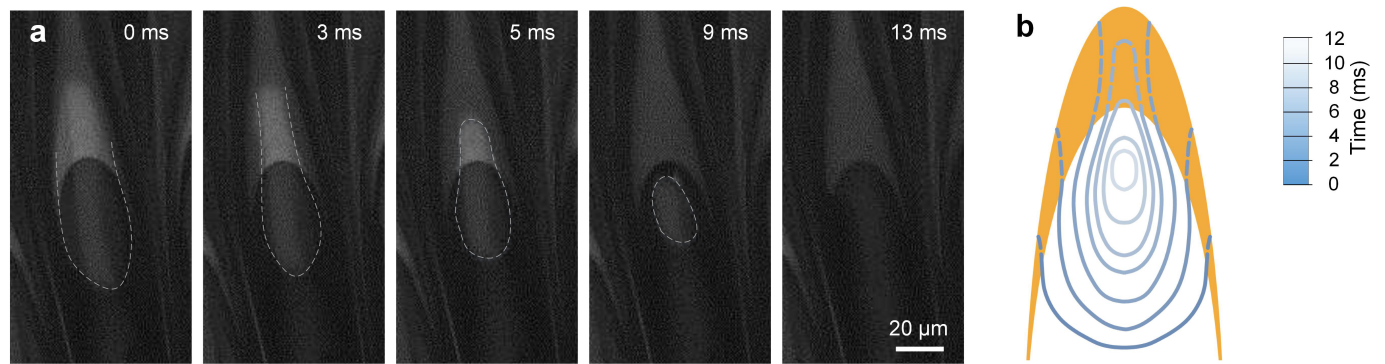


Extended Data Figure 1 | The peristome surface and the microcavity. **a**, SEM image of a cross-section of the peristome, showing that two-order, regular radial ridges align perfectly to form regular two-order parallel hierarchical microgrooves. **b**, SEM image showing a side view of the

vertical section of a microcavity. The image reveals the microcavity's geometry in the plane perpendicular to the peristome surface, and shows that the microcavity has a sharp edge with an angle, ϕ , of approximately 2° to 8° .

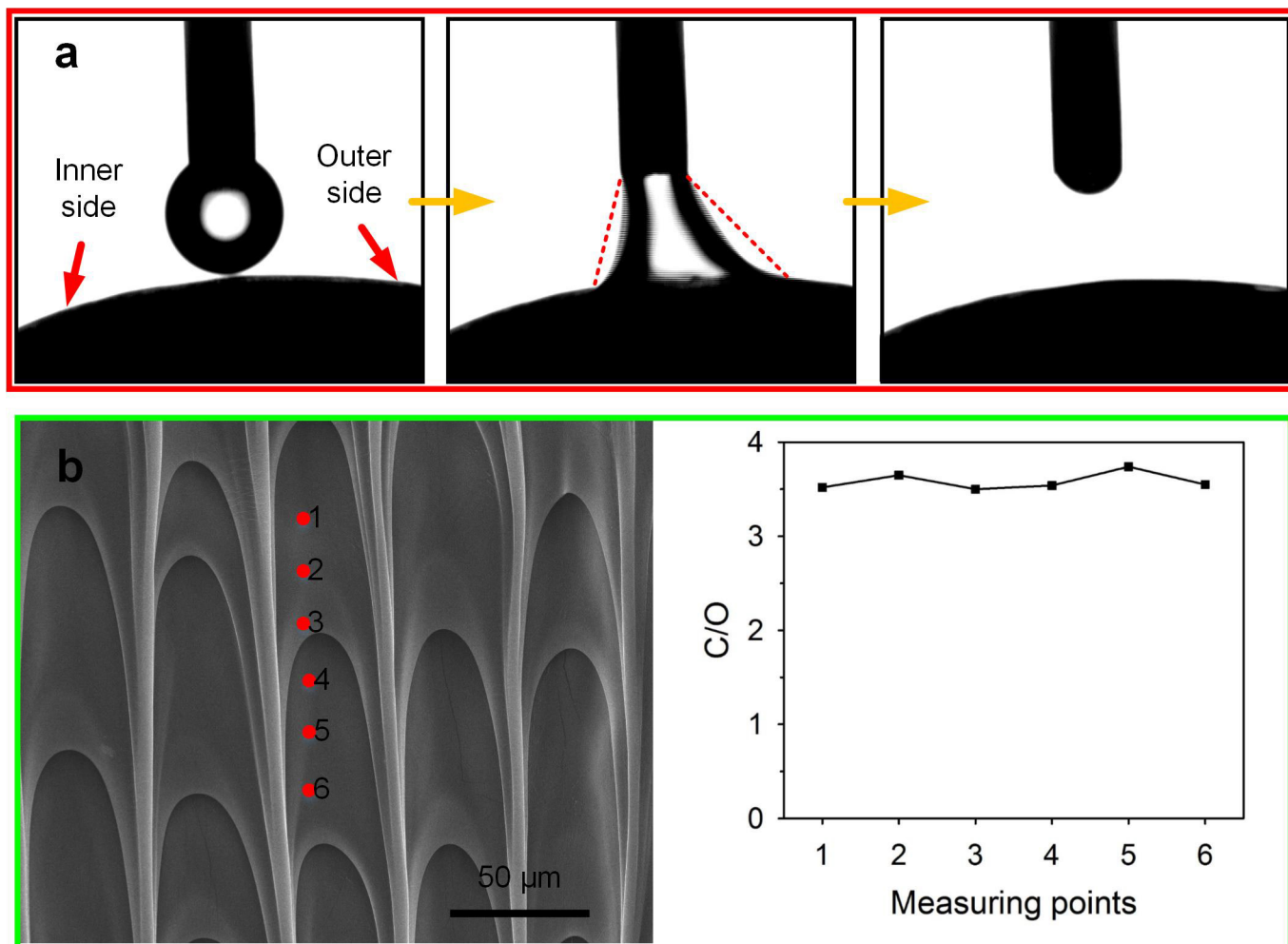


Extended Data Figure 2 | Cross-sections along the wedge corner of a single microcavity. a–h, Sequential cross-sectional images of the microcavity (moving deeper and deeper into the microcavity) show that it has a gradient wedge corner, with an opening angle that varies from roughly 90.5° to 28.2°.



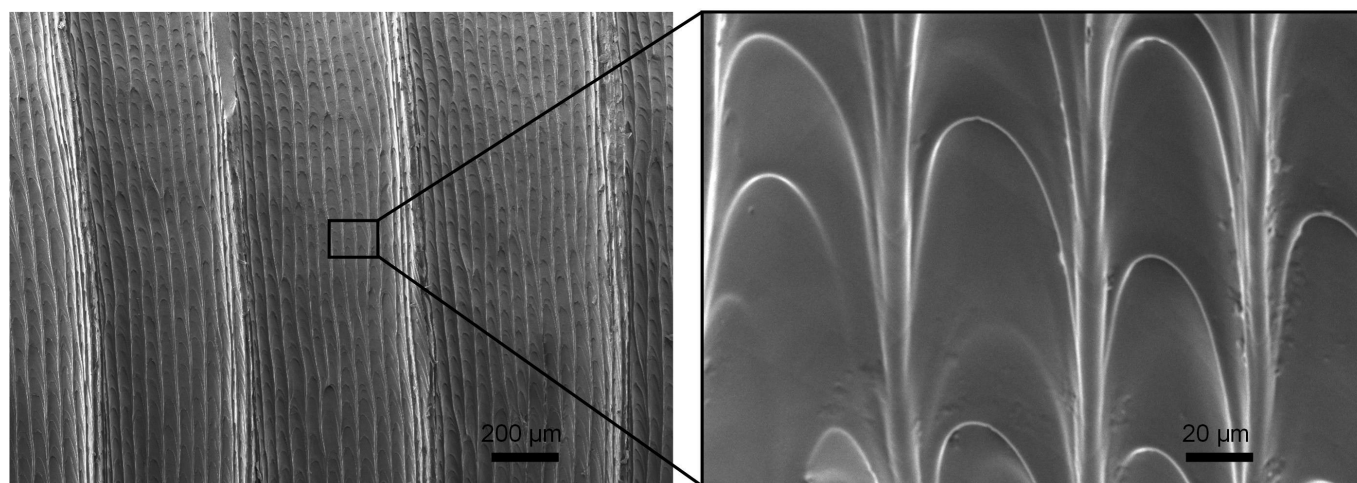
Extended Data Figure 3 | How a single microcavity fills with water. **a**, High-speed digital images of water filling a single duck-billed microcavity show that the water rapidly rises along the wedge corner

and squeezes air out of the microcavity, rising into the cavity to fill it completely. **b**, The water-filling process; blue lines correspond to the water boundary change with time.

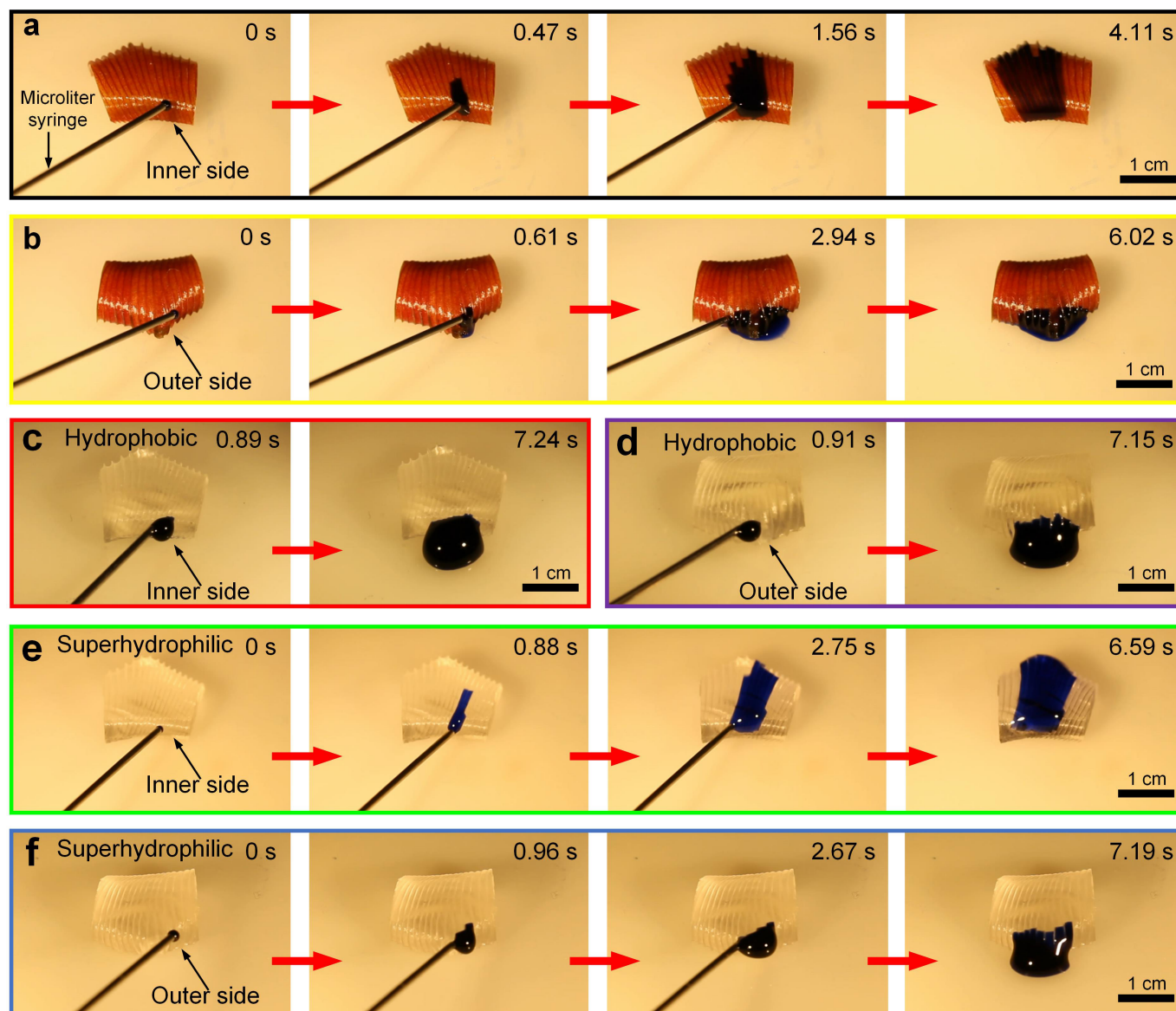


Extended Data Figure 4 | Wettability and chemical composition of the peristome. **a**, A water droplet (about 4 μl ; white circle) is dropped from a microsyringe and spreads quickly without a significant contact angle, indicating that the peristome is superhydrophilic. The water spreading also shows substantial directionality from the peristome's inner side to the

outer side. **b**, EDS measuring points along the second-order microgrooves, and associated carbon/oxygen (C/O) values. These six values are almost the same as each other, indicating no notable chemical gradient along the microgrooves.

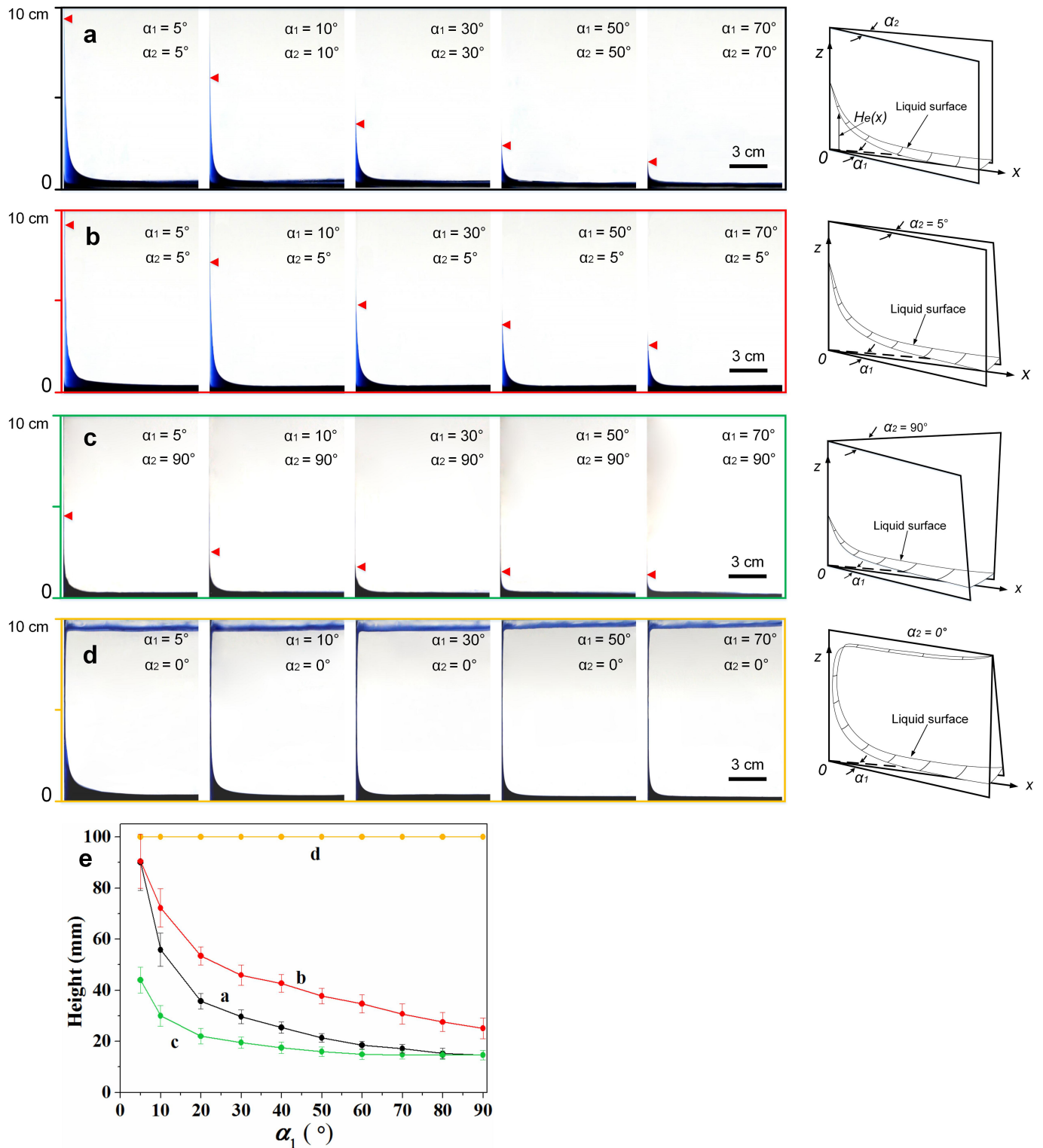


Extended Data Figure 5 | SEM images of PDMS replicas of the peristome. Like the natural peristome surface, the surfaces of the replicas also possess two-order microgrooves (left); moreover, duck-billed microcavities (right) are distributed along the second-order microgrooves.



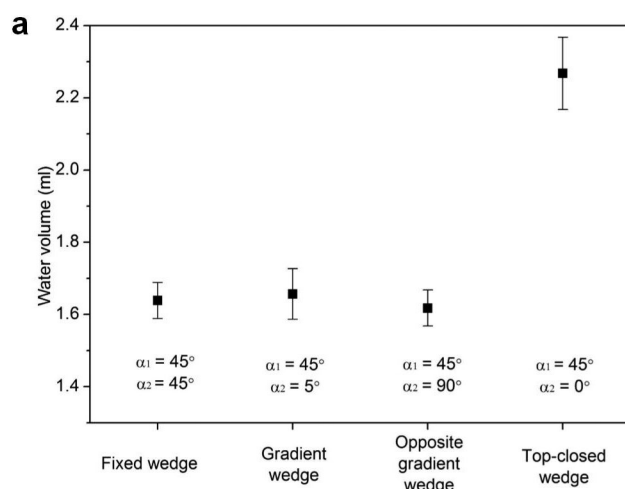
Extended Data Figure 6 | Directional water transport on natural peristome surfaces and PDMS replicas. **a, b**, The natural peristome. **a**, Water transport occurred when the water droplet made contact with the inner side. **b**, Water could not be transported when it made contact with the outer side. Thus, the natural peristome has a capacity for

directional water transport. **c, d**, A hydrophobic replica. Water could not be transported from either the inner (**c**) or the outer (**d**) side. **e, f**, A superhydrophilic replica. **e**, Water was transported from the inner side to the outer side. **f**, Water could not be transported from the outer side to the inner side.

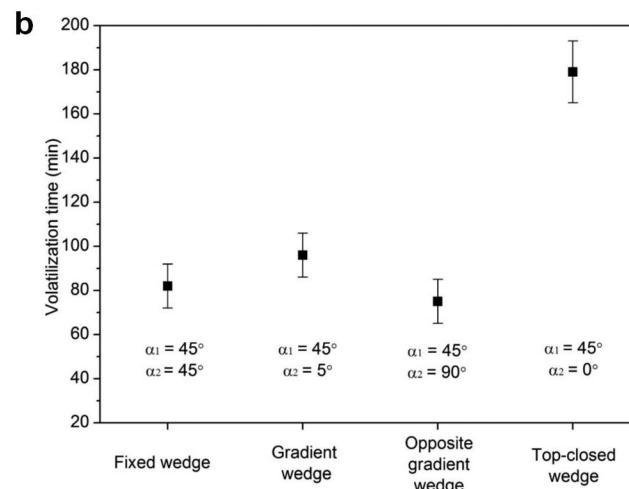


Extended Data Figure 7 | Experimental studies of water rise at various wedge corners. **a**, Capillary rise in a fixed wedge, in which the bottom opening angle, α_1 , equals the top opening angle, α_2 . The water rise decreases as the opening angle increases. **b**, Capillary rise in a gradient wedge, with $\alpha_2 = 5^\circ$, and $\alpha_1 > \alpha_2$; here, water rise is higher than that in the fixed wedge. **c**, Capillary rise in a wedge with the opposite gradient to that in panel **b**; $\alpha_2 = 90^\circ$, $\alpha_1 < \alpha_2$. Here, the water rise was the lowest of that seen in these four experiments. The red triangles in **a–c** show the height of the water rise. **d**, Capillary rise in a top-closed wedge, with $\alpha_2 = 0^\circ$. The water can rise along the vertical wedge corner to fill the horizontally

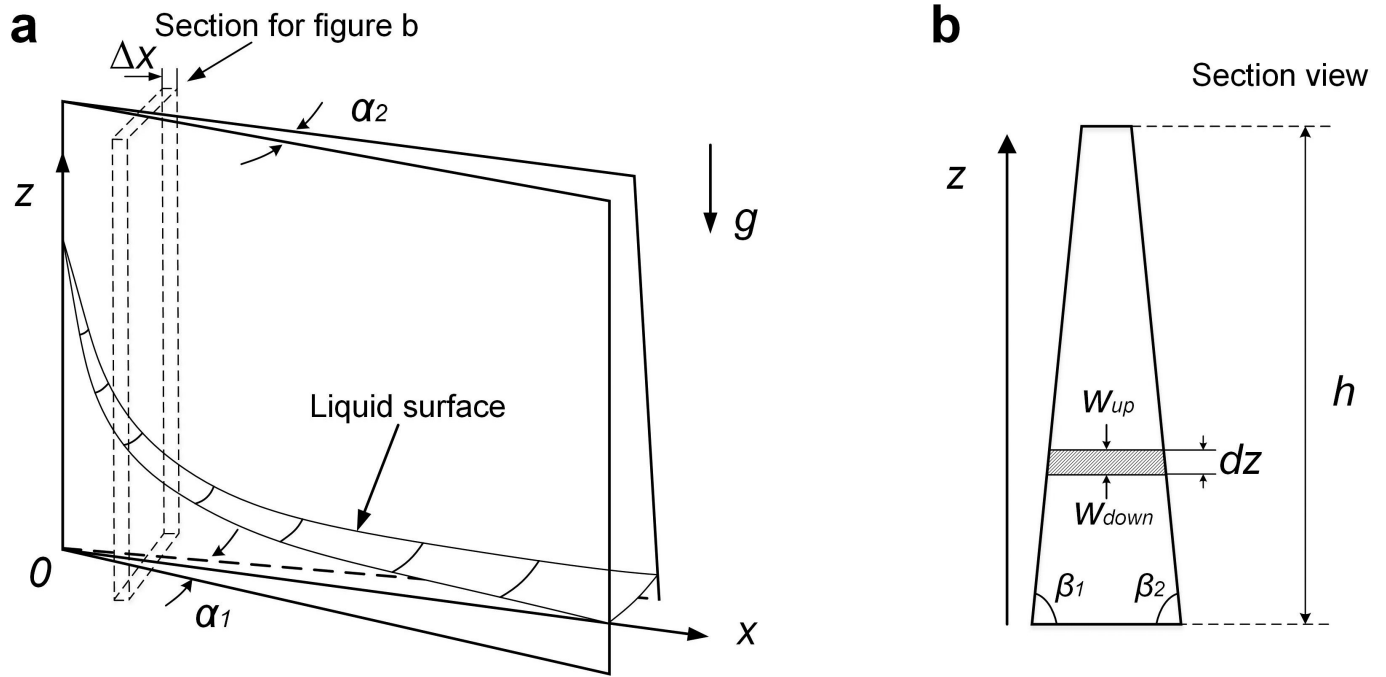
top-closed wedge independently of α_1 . Diagrams showing the water rise are at the right of **a–d**. The heights were limited to 100 mm because of the restricted sizes of the available material. **e**, The experimental water-rise results show that the water rise generally decreases with increasing α_1 . Water rise in the gradient wedge (red line) is higher than that in the fixed wedge (black line); the top-closed wedge (yellow line) shows additional horizontal water transport. Water rise was the lowest in the opposite gradient wedge (green line). Error bars indicate standard deviations from at least five independent measurements.



Extended Data Figure 8 | Volume of rise water and water-retention time for the four types of wedge corner. **a**, Volume of rise water in the four types of wedge—fixed wedge, gradient wedge, opposite gradient wedge, and top-closed wedge. The volume of rise water for the top-closed wedge is the largest, and is about 40% greater than that for the fixed wedge.



b, Water-retention time in the four types of wedge. Water-retention time for the top-closed wedge is the longest, and is about twice that for the fixed wedge. Error bars indicate standard deviations from at least five independent measurements.



Extended Data Figure 9 | Capillary rise in the gradient wedge, showing relevant variables. **a**, Liquid rise in the gradient wedge corner. **b**, Section view of the liquid rise in panel **a**. α_1 , bottom opening angle of a wedge; α_2 , top opening angle; g , gravitational constant; β_1 and β_2 , angle between

the plate and the horizontal plane ($\beta_1 = \beta_2$); h , height of the intersecting plates; w_{up} , up width of selected small liquid element; w_{down} , down width; dz , thickness of the small liquid element.

Nineteen-step total synthesis of (+)-phorbol

Shuhei Kawamura¹, Hang Chu¹, Jakob Felding² & Phil S. Baran¹

Phorbol, the flagship member of the tigliane diterpene family, has been known for over 80 years and has attracted attention from many chemists and biologists owing to its intriguing chemical structure and the medicinal potential of phorbol esters¹. Access to useful quantities of phorbol and related analogues has relied on isolation from natural sources and semisynthesis. Despite efforts spanning 40 years, chemical synthesis has been unable to compete with these strategies, owing to its complexity and unusual placement of oxygen atoms. Purely synthetic enantiopure phorbol has remained elusive, and biological synthesis has not led to even the simplest members of this terpene family. Recently, the chemical syntheses of eudesmanes², germacrenes³, taxanes^{4,5} and ingenanes^{6–8} have all benefited from a strategy inspired by the logic of two-phase terpene biosynthesis in which powerful C–C bond constructions and C–H bond oxidations go hand in hand. Here we implement a two-phase terpene synthesis strategy to achieve enantiospecific total synthesis of (+)-phorbol in only 19 steps from the abundant monoterpene (+)-3-carene. The purpose of this synthesis route is not to displace isolation or semisynthesis as a means of generating the natural product *per se*, but rather to enable access to analogues containing unique placements of oxygen atoms that are otherwise inaccessible.

Tigliane diterpenes have been viewed as promising leads in many different medicinal applications including immunomodulatory⁹, anti-viral¹⁰ and anti-cancer¹¹ applications. The most progressed compound in this class of natural products is phorbol 12-myristate 13-acetate, which is currently in phase II clinical trials for treatment of acute myeloid leukaemia. Subtle perturbations of their structures can have a marked effect on their biological profiles, perhaps as a result of differing protein kinase C subtype selectivity^{9,10}. As such, phorbol esters and related terpenes have been a focus of natural products research¹. Our interest in this family stemmed from the anti-cancer effects of certain phorbol analogues whose access was limited. Numerous eudesmane family members have previously been accessed in a concise way via a strategy for the chemical synthesis of complex terpenes that follows the underlying logic of biosynthesis². This strategy has subsequently been successfully applied to germacrenes³, complex taxanes^{4,5} and ingenanes^{6–8}. In particular, the two-phase synthesis of (–)-ingenol (2, Fig. 1) is only 14 steps from (+)-3-carene (6) and has enabled access to a variety of analogues that were not accessible by any other means^{6–8}. Because ingenanes (4) and tiglianes (3) are closely related in nature, we hypothesized that intermediate 5, which is available in quantities of more than 100 g, might serve as an ideal starting point for a short route to 1. The execution of this plan depended on the invention of a simple solution to the challenge of incorporating the C12 and C13 oxygen atoms of 1; the difficulty in installing this functionality is well known in the field of organic chemistry. So far, only two total syntheses^{12–14} and two formal syntheses^{15,16} of 1 have been reported in 40–52 steps (see Supplementary Information for a summary)^{12–16}. In these studies, an α -oxygenated enone was prepared from a ketone and subsequently cyclopropanated to build in the C13 oxidation via a six-step sequence. Many other efforts towards 1 have also been reported^{17,18} (for a full listing of the 36 papers in this area, see Supplementary Information).

The route described in refs 17 and 18 is particularly instructive and illustrates the inherent challenge of using (+)-3-carene (6) as a starting material to (+)-phorbol (1): 38 steps were required to reach a phorbol analogue lacking the C11 methyl and C13 oxygen from 6 (refs 17, 18). (Here we define a reaction step as one in which a substrate is converted to a product in a single reaction flask (irrespective of the number of transformations) without intermediate workup or purification.) Our studies, outlined below, culminated in a two-phase, 19-step synthesis of 1 that solves these problems in a concise way.

The synthesis commenced with intermediate 5 (Fig. 2), provided in large quantities using a previously described route⁶. Installation of the C4 oxygen was accomplished using a Mukaiyama hydration and *in situ* silyl group installation to furnish 7 in 70% yield (gram scale). At this juncture, we pursued installation of the C12 oxygen atom, which required the selective oxidation of a methylene position in the presence of an enone, six tertiary C–H bonds and two other competing methylene sites. To choose the proper oxidant for this transformation, we analysed the structure computationally (Fig. 3a) and by

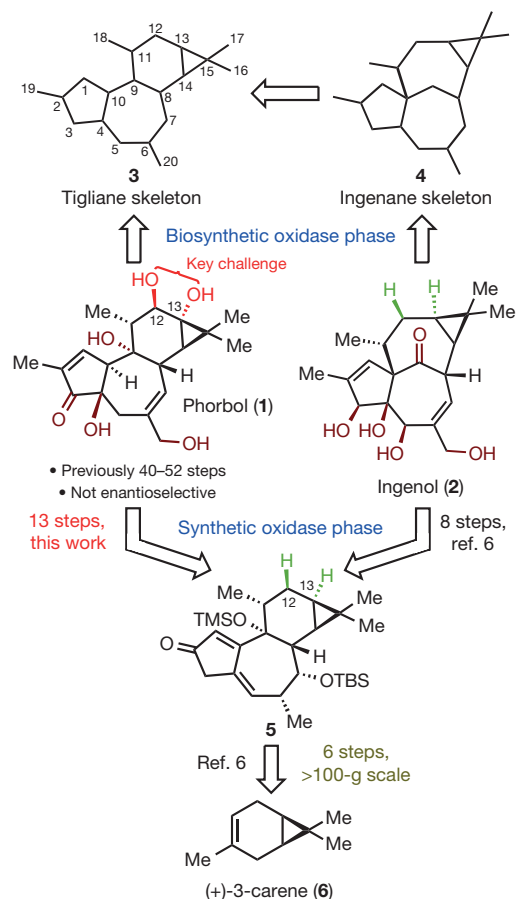


Figure 1 | A two-phase approach to ingenanes and tiglianes enables a concise approach to the phorbol structure.

¹Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA. ²Front End Innovation, LEO Pharma A/S Industriparken 55, 2750 Ballerup, Denmark.

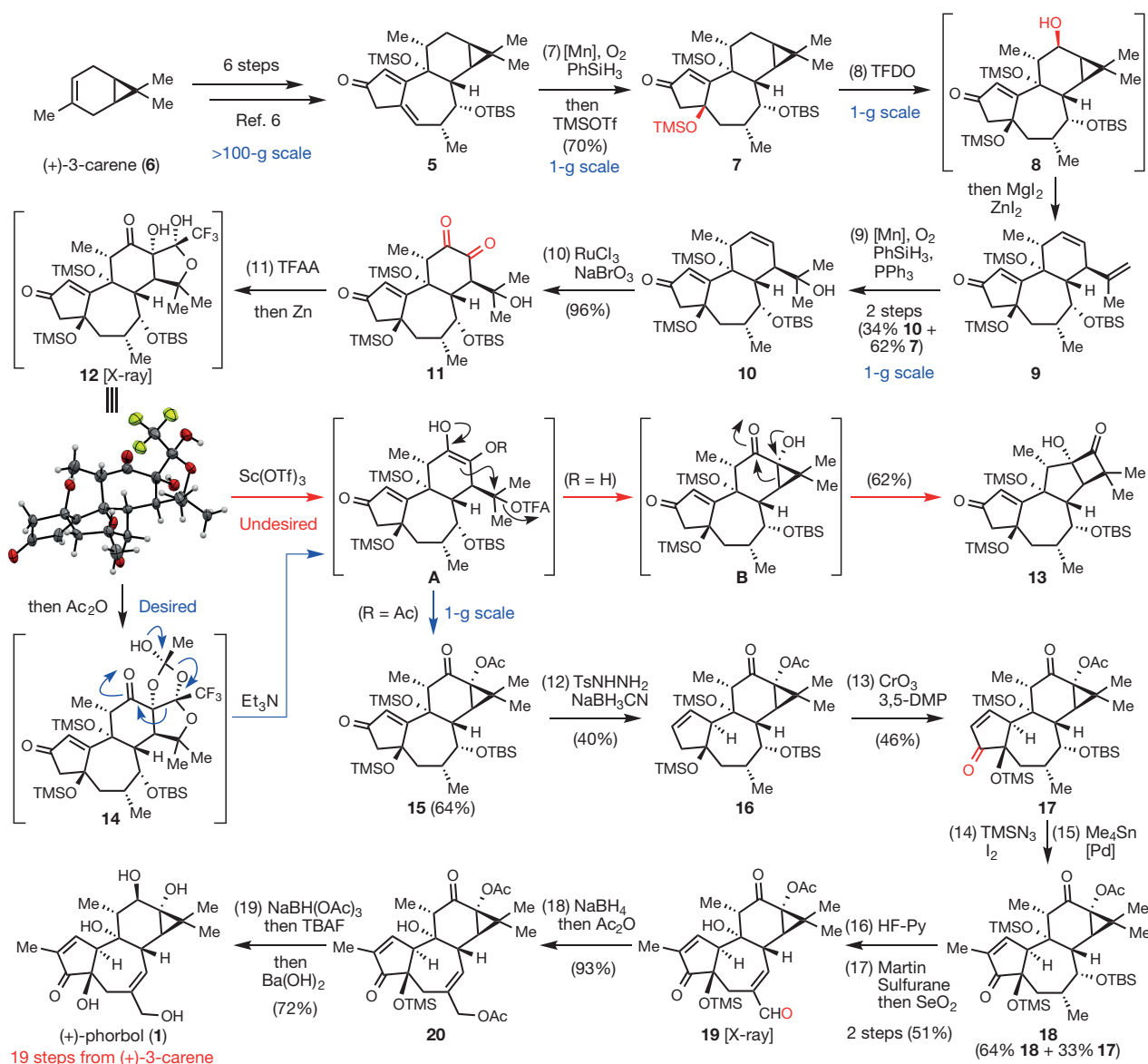


Figure 2 | 19-step total synthesis of 1. Reagents and conditions are as follows. (7) Mn(acac)₃, PhSiH₃, O₂, EtOH, then TMSOTf, Et₃N, CH₂Cl₂, 0 °C, 70%. (8) TFDO, CH₂Cl₂, 0 °C, then ZnI₂, MgI₂, Et₂O. (9) Mn(acac)₂, PhSiH₃, O₂, PPh₃, EtOH, 2 steps (34% **10** + 62% **7**). (10) RuCl₃, NaBrO₃, NaHCO₃, EtOAc, CH₃CN, H₂O, 96%. (11) TFAA, DMAP, CH₂Cl₂, 0 °C, then Zn, AcOH, CH₂Cl₂, then Ac₂O, DMAP, CH₂Cl₂, 0 °C, then Et₃N, DMF, 60 °C, 64%. (12) TsNHNH₂, MeOH, reflux, then NaBH₃CN, AcOH,

reflux, 40%. (13) CrO₃, 3,5-DMP, CH₂Cl₂, 0 °C to room temperature, 46%. (14) TMSN₃, DCE, 70 °C, then I₂, DCE, pyridine, 70 °C. (15) Me₄Sn, PdCl₂(PhCN)₂, AsPh₃, CuI, NMP, 80 °C, 2 steps (64% **18** + 33% **17**). (16) HF·pyridine, THF, 0 °C. (17) Martin Sulfurane, DCE, 60 °C, then SeO₂, benzene, 80 °C, 2 steps 51%. (18) NaBH₄, MeOH, −40 °C, then Ac₂O, DMAP, CH₂Cl₂, 93%. (19) NaBH(OAc)₃, benzene, reflux, then TBAF, THF, 0 °C, then Ba(OH)₂, MeOH, 72%.

inference of innate reactivity using NMR (Fig. 3b). Thus, we predicted the pseudo-equatorial C–H bond at C12 to be the most reactive on the basis of the following considerations¹⁹: (1) steric shielding of the C6, C7, C8 and C11 positions would decrease their rate of oxidation; (2) the higher *s*-character of the tertiary cyclopropane C–H bonds (C13/14) makes them very difficult to oxidize; (3) of the remaining carbon centres, ¹³C NMR indicated that C12 is the most ‘nucleophilic’; (4) hyperconjugation from the π-like cyclopropane system should facilitate oxidation of the pseudo-equatorial C–H bond on C12; and (5) strain-release might, to a small extent, accelerate such an oxidation²⁰. Therefore, we chose the small, reactive, electrophilic oxidant methyl(trifluoromethyl)dioxirane (TFDO), owing to its straightforward preparation and success in other challenging methylene oxidations²¹. As predicted, TFDO cleanly achieved C–H oxidation at C12 to deliver the intermediate cyclopropyl carbinol intermediate **8** (single diastereomer), which could be isolated and fully characterized on a gram scale. In practice, however, it was directly treated in the

same flask with MgI₂/ZnI₂ to elicit a dehydrative ring opening²² to furnish diene **9** along with unreacted enone **7**. We did not observe the anticipated tertiary iodide product, presumably owing to rapid elimination to form a diene system. The mixture of **7** and **9** was directly subjected to another Mukaiyama hydration to furnish the tertiary carbinol **10** in 34% isolated yield from **7** along with 62% recovered **7**. The addition of PPh₃ to this type of reaction has not previously been reported and was found to be essential to prevent the reactive intermediate peroxide adduct from generating an unwanted by-product (see Supplementary Information for details).

With the cyclopropane ruptured, the next step was to install the C12/13 oxygenation pattern along with reclosure of the cyclopropane. This manoeuvre has no precedent, and needed to be accomplished rapidly and in a scalable fashion. To this end, chemoselective oxidation of the C12/13 olefin to diketone **11** took place in near quantitative yield using Ru/NaBrO₃. To reform the *gem*-dimethylcyclopropane moiety and access the correct oxidation state, a two-electron reduction

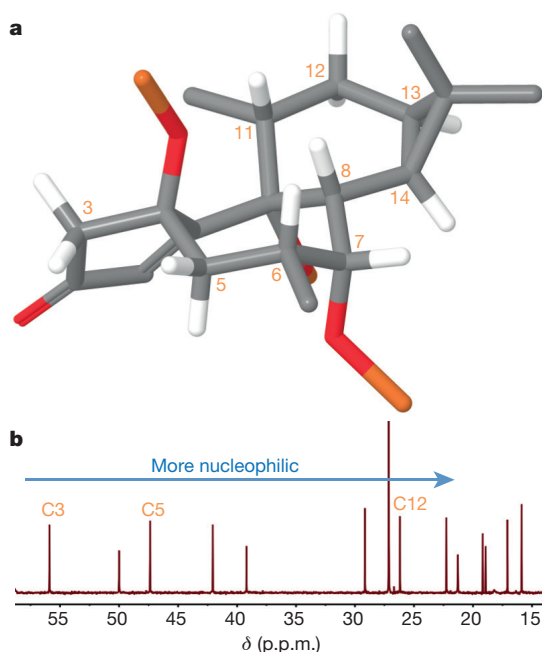


Figure 3 | Predicting the site of C–H functionalization on intermediate 7. **a**, Structural analysis of intermediate 7 calculated using MacroModel 10.8 (Schrödinger LLC; <http://www.schrodinger.com/citations/41/11/1/>). Grey, carbon; white, hydrogen; red, oxygen; orange, silicon. **b**, ^{13}C NMR can be used to predict the most nucleophilic methylene C–H bond after taking into account the role of other factors (see text).

of the diketone to an enediol would need to take place followed by intramolecular cyclization. Whereas attempts at acetylation of the hindered tertiary alcohol were fruitless and mesylation led to rapid elimination (as with the putative iodide intermediate in **8** \rightarrow **9**), treatment with trifluoroacetic anhydride smoothly led to an intermediate trifluoroacetate. This ester could be treated in the same flask with Zn to deliver an intermediate (**12**) whose structure was confirmed with the aid of X-ray crystallography. Although the desired reduction took place, it was accompanied by the formation of an additional, unwanted and bizarre C–C bond between C13 and the carbonyl group of the trifluoroester. Although aldol additions of enediols to aldehydes²³ and ketones²⁴ are known, there is no precedent for such reactivity with esters. We reasoned that this product originated from a highly reactive enediol intermediate (such as **A**; R = H) that might be regenerated under equilibrating conditions. In support of this hypothesis was the empirical observation that conversion of **11** to **12** initially resulted in the formation of an approximately 1:1 mixture of **12** along with an isomeric species tentatively assigned as the diastereomeric aldol adduct. Over 24 h this mixture equilibrated to **12** exclusively, perhaps owing to stabilizing intramolecular hydrogen bonding interactions. We evaluated various sets of conditions with the aim of reintroducing the enediol (retro aldol) in the hope that concomitant irreversible cyclopropane formation (via attack at C15) would take place. However, treatment of **12** with catalytic $\text{Sc}(\text{OTf})_3$ in DMF led instead to the cyclobutanone **13**. Although clearly undesired, this was encouraging because **13** must have been derived from the desired hydroxy-cyclopropane **B** via a 1,2-shift, which in turn was produced via the enediol intermediate **A** (R = H). Indeed, the formation of cyclobutanones such as **13** from hydroxycyclopropanes has a precedent²⁵. To circumvent this undesired pathway, intermediate **12** was converted to hemiothoester **14** by simple treatment with Ac_2O followed by exposure to Et_3N and gentle heating to 60°C to deliver **15**. The single flask conversion of **11** to **15** can be conducted on a gram scale, and presumably succeeds owing to the *in situ* formation of intermediate **A** (R = Ac), which prevents 1,2-shift after cyclopropanation. We believe that the conversion of **A** to the

desired cyclopropane is concerted because no elimination products were observed (see above). This reaction cascade consists of thirteen discreet events occurring in the same flask, forming two new C–C bonds, three new ring systems and three C–O bonds, along with the cleavage of one C–C bond, two ring systems and two C–O bonds. It unravels a complex polycyclic network to a kinetic endpoint using simple reagents thereby overcoming the greatest challenge posed by **1**.

The final eight steps of the synthesis served to install two additional oxygen atoms, one methyl group and the proper C12 and C10 stereochemistry. Attempts to install the key C10 stereocentre by way of conjugate reduction were met with failure despite extensive experimentation (see Supplementary Information for details). Thus, to achieve allylic transposition of **15**, the weakest point of the synthesis from an efficiency standpoint, we used TsNHNH_2 under reductive conditions, which led to **16** in 40% yield. Allylic oxidation with CrO_3 (ref. 26) delivered enone **17** in 46% yield. A two-step methyl group introduction was accomplished by formation of an α -iodoenone (TMSN_3 , I_2)²⁷ followed by a Stille coupling²⁸, which afforded **18** in 64% overall yield along with 33% recovered **17**. Selective removal of the C7 and C9 silyl groups with the HF · pyridine complex led to a diol that could be selectively dehydrated using the Martin Sulfurane reagent to an unstable olefin that was directly oxidized using SeO_2 to aldehyde **19** in 51% overall yield. The selective cleavage of the C7 and C9 silyl groups could stem from the inductive effect of the C3 carbonyl group, which could make the C4 silyl group less reactive to acids. X-ray crystallographic analysis of **19** confirmed the structural assignments made thus far. To complete the synthesis, **19** was reduced to the allylic alcohol and shielded as the acetate ester. Attempts to directly reduce the C20 aldehyde and the C12 ketone in **19** were unsuccessful, owing to the rapid rate of the C3 enone carbonyl reduction in the presence of the C20 free primary alcohol (the C20 hydroxy group is actually proximate to C3 as judged by molecular models). Stereoselective ketone reduction using $\text{NaBH}(\text{OAc})_3$ followed by successive treatment with fluoride and hydroxide sources (TBAF and $\text{Ba}(\text{OH})_2$, respectively) led to optically pure (+)-**1** (synthetic, $[\alpha]_D^{26}$ 105.37 (*c* 0.23, CH_3OH); natural, $[\alpha]_D^{27}$ 104.25 (*c* 0.20, CH_3OH)) in 72% isolated yield (15 mg prepared by this route).

The concise route to **1** is enabled by a fundamentally different retrosynthetic approach to terpene synthesis in the laboratory, rather than by focusing on the invention of a new synthetic method; the newest method used in this synthesis is a Stille coupling, invented in the 1980s²⁸. Although the development of new synthetic methodologies are necessary to push the field of organic chemistry forward, this work emphasizes the equal importance of strategy design in the total synthesis of complex natural products²⁹. By holistically mimicking the biosynthesis of ingenol (**2**) in the laboratory^{6,7}, we exploited the interrelationship between ingenanes and tiglanes by using the same building block for both synthetic routes. Because two-phase terpene synthesis builds the carbon skeleton with a strategically planned redox state, the majority of steps can focus on installing key oxidations rather than repositioning functional groups and C–C bonds. Access to phorbol (**1**) requires only five additional steps compared to ingenol (**2**), owing to the presence of two additional oxygen atoms (C12/13) placed at particularly challenging locations on the carbon skeleton. Successful installation of these oxygenations and two others directly or indirectly relied on the application of C–H functionalization logic to pinpoint both the location and sequence of reactions^{19,30}. This artificial oxidase phase will enable the synthesis of analogues containing deep-seated modifications in the same fashion that has recently been reported for ingenol (**2**)⁸. Although the route, in its current concise form, cannot compete with isolation for the procurement of large quantities of phorbol (**1**), we argue that it could be adapted to become truly scalable if the need arose. Rather, the purpose of this synthesis is to enable rapid access to new tiglane family members with exciting bioactivity that are either difficult or impossible to access through isolation or semisynthesis.

Received 4 December 2015; accepted 20 January 2016.

Published online 23 March 2016.

- Wang, H.-B., Wang, X.-Y., Liu, L.-P., Qin, G.-W. & Kang, T.-G. Tigliane diterpenoids from the euphorbiaceae and thymelaeaceae families. *Chem. Rev.* **115**, 2975–3011 (2015).
- Chen, K. & Baran, P. S. Total synthesis of eudesmane terpenes by site-selective C–H oxidations. *Nature* **459**, 824–828 (2009).
- Foo, K. *et al.* Scalable, enantioselective synthesis of germacrenes and related sesquiterpenes inspired by terpene cyclase phase logic. *Angew. Chem. Int. Ed.* **51**, 11491–11495 (2012).
- Mendoza, A., Ishihara, Y. & Baran, P. S. Scalable enantioselective total synthesis of taxanes. *Nature Chem.* **4**, 21–25 (2012).
- Wilde, N. C., Isomura, M., Mendoza, A. & Baran, P. S. Two-phase synthesis of (–)-taxuyunnanin D. *J. Am. Chem. Soc.* **136**, 4909–4912 (2014).
- Jørgensen, L. *et al.* 14-step synthesis of (–)-ingenol from (+)-3-carene. *Science* **341**, 878–882 (2013).
- McKerrall, S. J., Jørgensen, L., Kuttruff, C. A., Ungeheuer, F. & Baran, P. S. Development of a concise synthesis of (–)-ingenol. *J. Am. Chem. Soc.* **136**, 5799–5810 (2014).
- Jin, Y. *et al.* C–H oxidation of ingenanes enables potent and selective protein kinase C isoform activation. *Angew. Chem. Int. Ed.* **54**, 14044–14048 (2015).
- Isakov, N. & Altman, A. Regulation of immune system cell functions by protein kinase C. *Front. Immunol.* **4**, 384 (2013).
- McKernan, L. N., Momjian, D. & Kulkosky, J. Protein kinase C: one pathway towards the eradication of latent HIV-1 reservoirs. *Adv. Virol.* **2012**, 805347 (2012).
- Mackay, H. J. & Twelves, C. J. Targeting the protein kinase C family: are we there yet? *Nature Rev. Cancer* **7**, 554–562 (2007); corrigendum **8**, doi:10.1038/nrc2350 (2008).
- Wender, P. A. *et al.* Studies on tumor promoters. 8. The synthesis of phorbol. *J. Am. Chem. Soc.* **111**, 8957–8958 (1989).
- Wender, P. A., Lee, H. Y., Wilhelm, R. S. & Williams, P. D. Studies on tumor promoters. 7. The synthesis of a potentially general precursor of the tiglianes, daphnanes, and ingenanes. *J. Am. Chem. Soc.* **111**, 8954–8957 (1989).
- Wender, P. A., Rice, K. D. & Schnute, M. E. The first formal asymmetric synthesis of phorbol. *J. Am. Chem. Soc.* **119**, 7897–7898 (1997).
- Wender, P. A. & McDonald, F. E. Studies on tumor promoters. 9. A second-generation synthesis of phorbol. *J. Am. Chem. Soc.* **112**, 4956–4958 (1990).
- Lee, K. & Cha, J. K. Formal synthesis of (+)-phorbol. *J. Am. Chem. Soc.* **123**, 5590–5591 (2001).
- Sugita, K., Shigeno, K., Neville, C. F., Sasai, H. & Shibasaki, M. Synthetic studies towards phorbols: synthesis of B or C ring substituted phorbol skeletons in the naturally occurring form. *Synlett* **1994**, 325–329 (1994).
- Sugita, K., Neville, C. F., Sodeoka, M., Sasai, H. & Shibasaki, M. Stereocontrolled syntheses of phorbol analogs and evaluation of their binding affinity to PKC. *Tetrahedr. Lett.* **36**, 1067–1070 (1995).
- Newhouse, T. & Baran, P. S. If C–H bonds could talk: selective C–H bond oxidation. *Angew. Chem. Int. Ed.* **50**, 3362–3374 (2011).
- Zou, L. *et al.* Enhanced reactivity in dioxirane C–H oxidations via strain release: a computational and experimental study. *J. Org. Chem.* **78**, 4037–4048 (2013).
- Michaudel, Q. *et al.* Improving physical properties via C–H oxidation: chemical and enzymatic approaches. *Angew. Chem. Int. Ed.* **53**, 12091–12096 (2014).
- McCormick, J. P. & Barton, D. L. Synthetic applications of metal halides. Conversion of cyclopropylmethanols into homoallylic halides. *J. Org. Chem.* **45**, 2566–2570 (1980).
- Miyoshi, N., Takeuchi, S. & Ohgo, Y. A facile synthesis of 2,3-dihydroxyketones from 1,2-diketones and aldehydes using samarium diiodide. *Chem. Lett.* **22**, 959–962 (1993).
- Krohn, K., Frese, P. & Flörke, U. Biomimetic synthesis of the racemic angucyclinones of the aquayamycin and WP 3688-2 Types. *Chemistry* **6**, 3887–3896 (2000).
- Bartsch, H. & Hecker, E. Zur chemie des phorbols, XIII. Über eine acyloinumlagerung des 12-desoxy-12-oxo-phorbol-13,20-diacetats. *Liebigs Ann. Chem.* **725**, 142–153 (1969).
- Salmond, W. G., Barta, M. A. & Havens, J. L. Allylic oxidation with 3,5-dimethylpyrazole. Chromium trioxide complex. Steroidal Δ^5 -7-ketones. *J. Org. Chem.* **43**, 2057–2059 (1978).
- Sha, C.-K. & Huang, S.-J. Synthesis of β -substituted α -iodocycloalkenones. *Tetrahedr. Lett.* **36**, 6927–6928 (1995).
- Stille, J. K. The palladium-catalyzed cross-coupling reactions of organotin reagents with organic electrophiles. *Angew. Chem. Int. Ed. Engl.* **25**, 508–524 (1986).
- Nicolaou, K. C. & Sorensen, E. J. *Classics in Total Synthesis: Targets, Strategies, Methods* 821 (Wiley, 1996).
- Gutekunst, W. R. & Baran, P. S. C–H functionalization logic in total synthesis. *Chem. Soc. Rev.* **40**, 1976–1991 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by LEO Pharma, the Uehara Memorial Foundation (postdoctoral fellowship to S.K.) and the National Institute of General Medical Sciences Grant GM-097444. We are especially grateful to S. Natarajan of KemXtree and his team for providing ample quantities of compound **5**. We thank D.-H. Huang and L. Pasternack for assistance with NMR spectroscopy, and A. L. Rheingold and C. E. Moore for X-ray crystallographic analysis.

Author Contributions S.K. and P.S.B. conceived this work; J.F. provided compound **5**; S.K., H.C. and P.S.B. designed the experiments and analysed that data; S.K. and H.C. conducted the experiments; S.K. performed the molecular mechanics calculations; and S.K. and P.S.B. wrote the manuscript.

Author Information Metrical parameters for the structures of **12** and **19** are available free of charge from the Cambridge Crystallographic Data Centre (CCDC) under reference numbers 1434376 and 1434377. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.S.B. (pbaran@scripps.edu).

Northern Hemisphere hydroclimate variability over the past twelve centuries

Fredrik Charpentier Ljungqvist^{1,2,3}, Paul J. Krusic^{4,5}, Hanna S. Sundqvist^{3,5}, Eduardo Zorita⁶, Gudrun Brattström^{3,7} & David Frank⁸

Accurate modelling and prediction of the local to continental-scale hydroclimate response to global warming is essential given the strong impact of hydroclimate on ecosystem functioning, crop yields, water resources, and economic security^{1–4}. However, uncertainty in hydroclimate projections remains large^{5–7}, in part due to the short length of instrumental measurements available with which to assess climate models. Here we present a spatial reconstruction of hydroclimate variability over the past twelve centuries across the Northern Hemisphere derived from a network of 196 at least millennium-long proxy records. We use this reconstruction to place recent hydrological changes^{8,9} and future precipitation scenarios^{7,10,11} in a long-term context of spatially resolved and temporally persistent hydroclimate patterns. We find a larger percentage of land area with relatively wetter conditions in the ninth to eleventh and the twentieth centuries, whereas drier conditions are more widespread between the twelfth and nineteenth centuries. Our reconstruction reveals that prominent seesaw patterns of alternating moisture regimes observed in instrumental data^{12–14} across the Mediterranean, western USA, and China have operated consistently over the past twelve centuries. Using an updated compilation of 128 temperature proxy records¹⁵, we assess the relationship between the reconstructed centennial-scale Northern Hemisphere hydroclimate and temperature variability. Even though dry and wet conditions occurred over extensive areas under both warm and cold climate regimes, a statistically significant co-variability of hydroclimate and temperature is evident for particular regions. We compare the reconstructed hydroclimate anomalies with coupled atmosphere–ocean general circulation model simulations and find reasonable agreement during pre-industrial times. However, the intensification of the twentieth-century-mean hydroclimate anomalies in the simulations, as compared to previous centuries, is not supported by our new multi-proxy reconstruction. This finding suggests that much work remains before we can model hydroclimate variability accurately, and highlights the importance of using palaeoclimate data to place recent and predicted hydroclimate changes in a millennium-long context^{16,17}.

Global warming is expected to strongly influence the hydrological cycle^{2–4,18,19}. Model simulations and theory suggest an intensification of the wet–dry contrast at lower latitudes⁵ and a widespread expansion of dry areas over land³. These projections have yet to be robustly detected in recent instrumental observations^{8,9}. Furthermore, regional changes in precipitation patterns and drought intensity, in response to global warming, are challenging to predict owing to the chaotic nature of hydroclimate variability, the low spatial coherence of precipitation, and fine-scale processes related to cloud formation and precipitation, resulting in the large spread among model simulations^{6,7}. From proxy evidence, we know that hydroclimate anomalies of larger amplitude and longer duration than found in modern instrumental observations have occurred in the past millennium: for example, the medieval

‘megadroughts’ in the western USA²⁰, and the monsoon failures in east Asia²¹ in the fifteenth–nineteenth centuries. This strongly suggests that the instrumental period is too short to capture the full range of natural hydroclimate variability. In this light, a comprehensive aggregation and assessment of millennium-long moisture-sensitive palaeoclimate proxy data are needed to improve our knowledge of past hydroclimate variability, provide a test bed for model simulations, and facilitate the detection of anthropogenic impacts on the hydrological cycle¹⁶.

Here, we analyse the spatio-temporal patterns of hydrological anomalies in an unprecedentedly large network of 196 hydroclimate proxy records from Northern Hemisphere land areas over the past twelve centuries. All are previously published as indicators of precipitation, drought, moisture balance, stream flow, lake level change, bog surface wetness or similar hydrological information derived from archives such as ice cores, marine sediments, lake sediments, speleothems, tree rings, and historical documentary data (Fig. 1a; Supplementary Table 1). All records span at least the past millennium, have at least two observations per century, and passed a screening procedure for geochronological dating control (Methods). Using the same selection and screening criteria, we updated and improved the temperature proxy network from ref. 15 (now 128 records), to investigate the co-variability between hydroclimate and temperature (Fig. 1b; Supplementary Table 2).

To develop a spatial reconstruction we standardized each proxy record from its original units (for example, $\delta^{18}\text{O}$, ring width, millimetres of precipitation) to an index, ranging from -2 to $+2$ standard deviations, of centennial variation (Methods). Anomalies exceeding -2 or $+2$ are considered outliers and truncated in order to minimize the influence of a few ‘extreme’ values. This index is interpreted as a measure of relative water availability, analogous to the Palmer Drought Severity Index (PDSI)^{20,21}, which permits comparisons of relative variations in moisture conditions between arid and moist regions.

This standardization is of particular importance given the different and potentially biased high- to low-frequency signal ratio in the various proxies. It essentially transforms the centennial-scale information in each proxy record to the same amplitude regardless of possible spectral biases (Methods). To extract a common spatial signal from the heterogeneous proxy data, the indices are spatially averaged over an estimated correlation decay length (latitudinally dependent) of centennial-scale hydroclimate and temperature variability, respectively, derived from model simulations and instrumental data (Methods; Extended Data Fig. 1). Unlike field reconstruction methods that rely upon teleconnections found in the instrumental period²², our technique yields a grid of hydrological anomalies that is independent of any spatial co-variability found in gridded instrumental observations¹⁵.

Our results reveal a general tendency towards a larger percentage of land area with relatively wetter conditions during the ninth–eleventh and twentieth centuries, whereas the twelfth–nineteenth centuries show larger areas with relatively drier conditions (Fig. 2). Centennial periods of reconstructed widespread dry conditions occurred under

¹Department of History, Stockholm University, SE-10691 Stockholm, Sweden. ²Centre for Medieval Studies, Stockholm University, SE-10691 Stockholm, Sweden. ³Bolin Centre for Climate Research, Stockholm University, SE-10691 Stockholm, Sweden. ⁴Navarino Environmental Observatory, GR-24001 Messinia, Greece. ⁵Department of Physical Geography, Stockholm University, SE-10691 Stockholm, Sweden. ⁶Helmholtz-Zentrum Geesthacht, Institute for Coastal Research, DE-21502 Geesthacht, Germany. ⁷Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden. ⁸Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland.

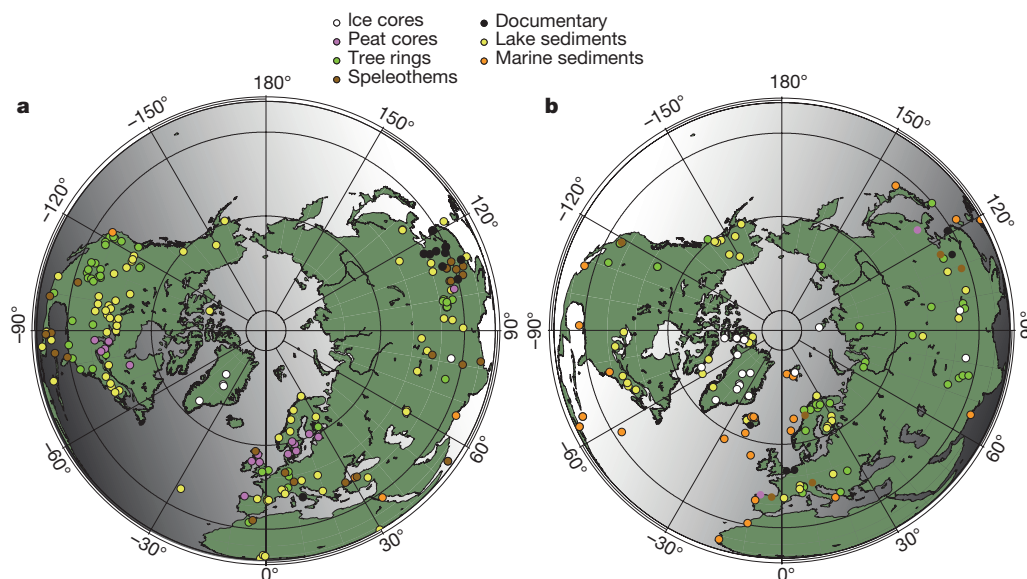


Figure 1 | Location map of the proxy records used in this study sorted by archive. a, Locations of the 196 hydrological proxy records for the past twelve centuries. **b,** Locations of the 128 temperature proxy records for the

past twelve centuries. All the records cover at least the period 1000–1899. We note that some proxies share the same geographical position but are based on different types of archives.

both warm and cold climate states (Fig. 3). Notably, the dry twelfth century (with 27% reconstructed hydroclimate anomalies < -0.5) is associated with positive forcing anomalies from high solar activity and moderate volcanic activity, whereas the equally dry fifteenth century (also with 27% reconstructed hydroclimate anomalies < -0.5) coincides with negative forcing anomalies from low solar forcing and several large volcanic eruptions²³. We identify inverse hydroclimate conditions (that is, hydroclimate dipoles) over geographically adjacent regions that persist on centennial timescales. This phenomenon has been reported in studies of instrumental data and some proxy time series for the same regions: the eastern and western Mediterranean¹²; the southwestern and northwestern USA¹³; eastern China and the Tibetan plateau¹⁴ (Fig. 2).

To assess the robustness of our gridded reconstruction we conducted a number of validation tests. Recognizing the potential for low-resolution, less well dated, proxies to introduce temporal biases in the reconstruction, we produced a decadal resolved reconstruction from an even more strictly screened subset of proxies (Methods). At the grid point level the median correlation between the full reconstruction and the decadal resolved reconstruction is $r = 0.97$ ($p < 0.01$). Additional sensitivity tests, for example, excluding one proxy type at a time, excluding or retaining records according to their seasonal signal, or their resolution (Methods; Extended Data Table 1), also reaffirm the robustness of the full reconstruction. Moreover, our reconstruction shows a relatively good correlation with tree-ring-derived reconstructions of PDSI, despite the differences between metrics and reconstruction seasonality (Methods; Extended Data Table 1)^{20,21}. We also compared our decadal reconstruction to decadal means of gridded instrumental precipitation data over the twentieth century. The median correlation is $r = 0.66$ ($p < 0.05$), but the short length of the instrumental data

precludes a full validation of the low-frequency behaviour in our hydroclimate reconstruction (Methods; Extended Data Figs 2 and 3).

In regions with sufficient overlap between the hydroclimate and temperature reconstructions, we observe a tendency for wetter conditions during warmer periods in northern Europe, the eastern Mediterranean, Greenland, northern China, northwestern North America and the southeastern USA (Extended Data Fig. 4c, d). Drier conditions during warmer periods are seen in the central Mediterranean, central and northeastern North America, southwestern USA, and southern China. For certain regions (for example, northern China, northwestern North America, northern Europe and Greenland) the patterns of covariance between reconstructed temperature and hydroclimate are qualitatively consistent with the predicted signature of hydroclimate changes under future global warming^{3,4,7,18,19}.

With our new spatial hydroclimate reconstruction we evaluated the centennial distributions of gridded hydrological and temperature anomalies (Fig. 3) against an identical extraction of precipitation and temperature anomalies from six state-of-the-art coupled atmosphere–ocean general circulation model simulations covering the period 850–2005 (Table 1; Fig. 2c; Extended Data Figs 4a, b and 5). Both the proxy and model data agree that the thirteenth century had the most extensive dry conditions in the pre-industrial period. We observe that the seventeenth century, the coldest Northern Hemisphere century of the past twelve centuries in both proxy and model data (Extended Data Fig. 6a, b)^{15,17}, shows hydrological conditions close to the long-term mean, with a lack of extremes (only 6% reconstructed hydroclimate anomalies exceeding ± 0.5).

Overall, large-scale agreement in the temporal evolution between proxy and model anomaly distribution characteristics is observed from the tenth to the eighteenth centuries for both hydroclimate and

Table 1 | Models used in this study from CMIP5, with simulations from 850 to 2005

Abbreviated model name	Model institute	Atmospheric resolution (longitudinal x latitudinal)	Forcing
bcc-csm1-1	Beijing Climate Center, China	2.81° × 2.81°	ref. 23
CCSM4	National Center for Atmospheric Research, US DOE/NSF, USA	1.25° × 0.9°	ref. 23
GISS-E2R	NASA Goddard Institute for Space Studies, USA	2.5° × 2°	ref. 23
HadCM3	Hadley Center, UK Met Office, UK	3.75° × 1.875°	ref. 23
IPSL-CM5A-LR	Institute Pierre Simon Laplace, France	3.75° × 1.89°	ref. 23
MPI-ESM-P	Max Planck Institute for Meteorology, Hamburg, Germany	1.875° × 1.875°	ref. 23

CMIP5, Climate Model Intercomparison project Phase 5. For references to the individual models and additional information, see Methods.

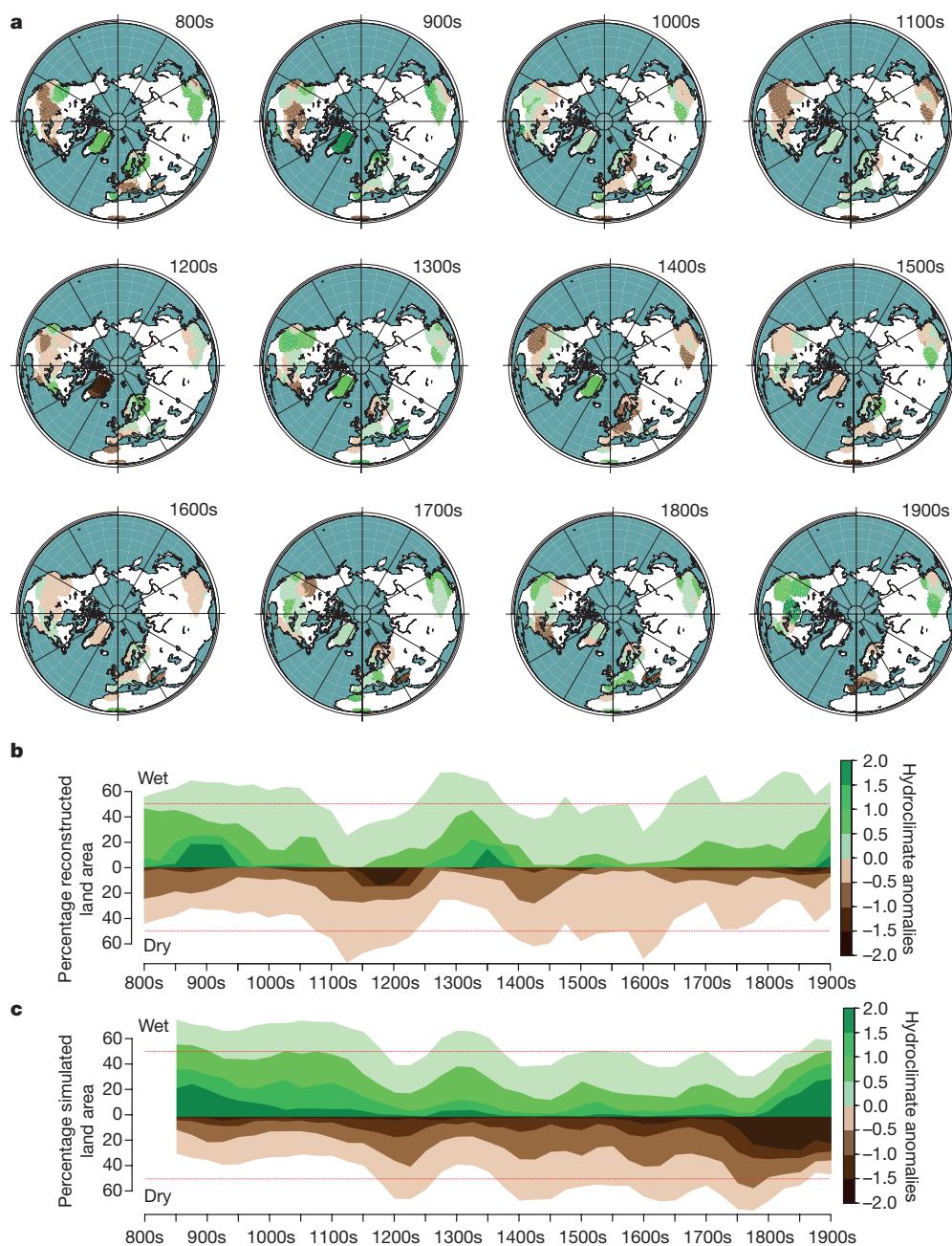


Figure 2 | Spatial-temporal distribution of gridded centennial hydrological proxy anomalies. **a**, Gridded, weighted, centennial hydroclimate proxy anomalies, derived from the proxy data shown in Fig. 1a, over land areas with at least three independent proxies within the estimated centennial correlation decay length for centennial-scale hydrological variability. **b**, Time series derived from the reconstructed gridded weighted hydroclimate anomalies, showing the fraction of land

area exceeding a given wetness or dryness threshold. The red horizontal bars denote the 50% levels. **c**, The same as **b** but for median-model anomaly values of annual precipitation. All anomalies are shown relative to the centennial mean and standard deviation over the eleventh to the nineteenth centuries. The colour scale is truncated at -2 and 2 and areas with insufficient proxy coverage to compute a gridded weighted mean value are left white.

temperature (Fig. 3). However, the similarities in reconstructed and simulated hydroclimate begin to break down in the nineteenth century, with outstanding disagreement of the centennial distribution characteristics in the twentieth century. The proxy data continue to show hydroclimate characteristics in the twentieth century similar to those in the pre-industrial era, whereas the models simulate a high percentage (36.8%) of the land areas to be in either an extremely dry (21.4% of anomalies < -1.5) or extremely wet (15.4% of anomalies > 1.5) state. In contrast to the simulations, the proxy data show more extreme hydroclimate anomalies during centuries other than the twentieth (Figs 2 and 3). The unprecedented intensification of wet and dry anomalies in the model simulations, presumably driven by anthropogenic forcing,

results in an exceptionally strong meridional gradient producing a wetter north and drier south (Fig. 2; Extended Data Fig. 7a–f) not found in the proxy data. This produces a distinct bimodal distribution of simulated hydroclimate (that is, wet and dry extremes), whereas the proxy data reveal a unimodal distribution both in the industrial and pre-industrial period (Fig. 3). Assessment of interdecadal hydroclimate trends and their variances, in proxy and model data, shows that the differences in the twentieth century between proxy and model data are mainly restricted to the centennial-scale mean anomaly. The model and proxy data agree that neither the interdecadal variability nor the decadal trends in the twentieth century are unprecedented (Supplementary Tables 5 and 6).

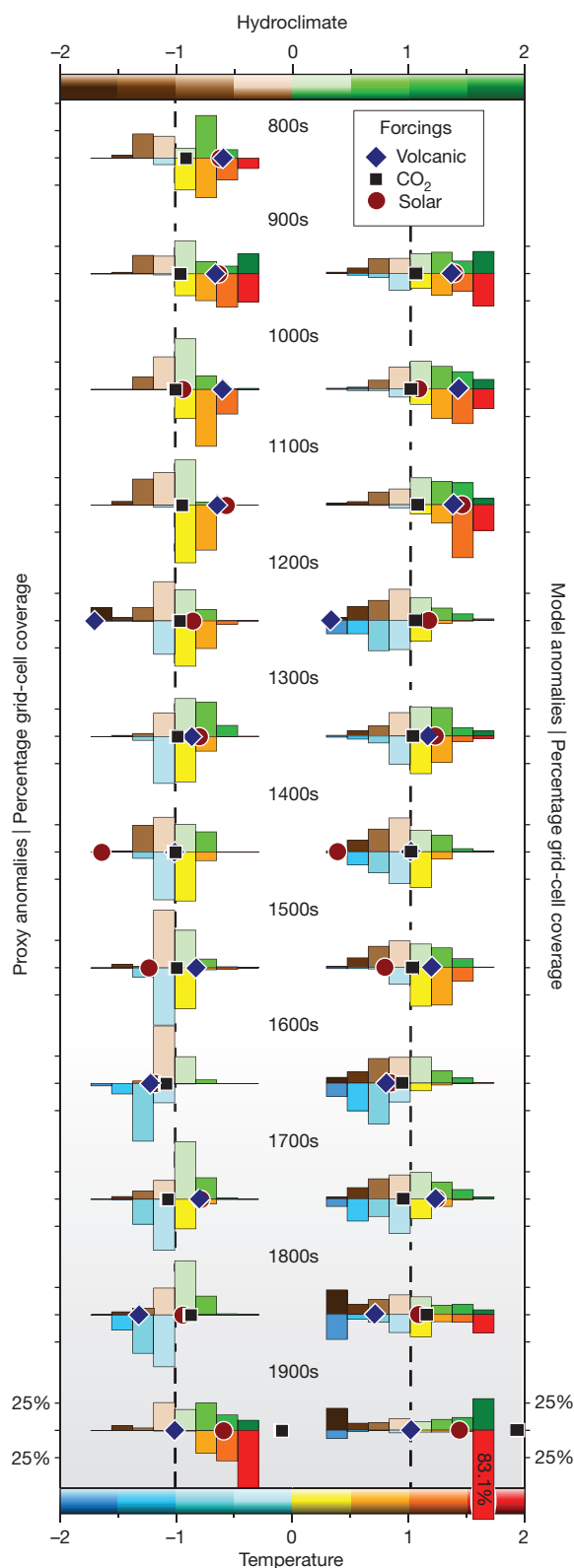


Figure 3 | Probability density function of reconstructed and simulated Northern Hemisphere hydroclimate and temperature anomalies. Reconstructed weighted-smoothed gridded proxy (left) and median-model (right) centennial anomaly values of Northern Hemisphere hydroclimate (top) and temperature (bottom), showing the percentages of Northern Hemisphere land area within defined threshold values ranging from -2 to $+2$ in increments of 0.5 standard normal deviates. Superimposed on each distribution are the corresponding centennial anomalies of solar, volcanic, and CO_2 forcing time-series (Methods).

A persistent link has been suggested between Northern Hemisphere thermal conditions and the displacement of the Intertropical Convergence Zone (ITCZ)²⁴. During warmer periods the ITCZ moves northward, changing the position of Northern Hemisphere storm tracks, monsoon activity and precipitation patterns²⁵. A hypothesized mechanism for these changes is an expansion of the Hadley circulation, bringing more moisture and latent energy northward. When we assess the reconstructed hydroclimate patterns during the reconstructed warmest (tenth) and coldest (seventeenth) pre-industrial centuries in the Northern Hemisphere, we find evidence for strong and weak, respectively, meridional gradients supporting the proposed dynamics. In our reconstruction, the twentieth-century manifestation of these dynamics cannot be seen to the extent reconstructed in the tenth century or in the twentieth century in model simulations (Fig. 2b, c; Extended Data Fig. 7). These findings could corroborate the suggestion by ref. 24 that ITCZ migration, owing to anthropogenic greenhouse gas forced warming and possibly tropospheric aerosols, may produce a hydroclimate regime that differs from past regimes. An alternative explanation is that the spatial signature of hydroclimate can be highly variable under comparable thermal conditions, making it difficult to attribute a particular external forcing to the observed patterns.

There are several possible explanations for the limited ability of climate models to simulate hydroclimate variability accurately: knowledge gaps in describing physical processes defining model sub-grid parameterizations^{6,26}, systematic absolute temperature biases in the models affecting evaporation and precipitation processes²⁷, and an underestimation of internal hydroclimate variability²⁸. We find only limited regional-scale reconstruction–model agreement (Fig. 2a–c and Extended Data Fig. 4a, b), as would be expected if Northern Hemisphere hydrological variability is driven more by internal dynamics than external forcing.

Our comparison of the distribution characteristics and large-scale temporal evolution of reconstructed and simulated centennial-scale hydroclimate reveals strong similarities up to the nineteenth century. However, the proxy evidence does not support the tendency in simulations for wet regions to become wetter and dry regions drier in a warmer climate^{29,30}. Furthermore, our hydroclimate reconstruction does not support a general unprecedented intensification of the hydrological cycle in the twentieth century, associated with both more extreme wet and dry conditions, as simulated by an ensemble of models. This finding is in line with recent analyses of instrumental data reporting limited evidence for an intensification of wet and dry anomalies under current global warming^{8,9,26}. Absence of evidence for this predicted hydroclimate response to higher temperatures suggests that either the warming in the twentieth century has not been strong enough for this fingerprint to emerge, or that the triggers for the mechanisms responsible should be revised. We conclude that a millennium-long perspective, as provided by multi-proxy evidence, is crucial to our ability to benchmark and evaluate^{16,17} hydroclimate simulations accurately.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 April 2015; accepted 6 February 2016.

1. D'Ondorio, P. & Bhattachan, A. Hydrologic variability in dryland regions: impacts on ecosystem dynamics and food security. *Phil. Trans. R. Soc. Lond. B* **367**, 3145–3157 (2012).
2. Field, C. B. et al. (eds) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).
3. Cook, B. I., Smerdon, J. E., Seager, R. & Coats, S. Global warming and 21st century drying. *Clim. Dyn.* **43**, 2607–2627 (2014).
4. Schewe, J. et al. Multi-model assessment of water scarcity under climate change. *Proc. Natl Acad. Sci. USA* **111**, 3245–3250 (2014).
5. Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006).

6. Stephens, G. L. *et al.* Dreary state of precipitation in global models. *J. Geophys. Res.* **115**, D24211 (2010).
7. Collins, M. *et al.* In *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 1029–1136 (Cambridge Univ. Press, 2013).
8. Sheffield, J., Wood, E. F. & Roderick, M. L. Little change in global drought over the past 60 years. *Nature* **491**, 435–438 (2012).
9. Greve, P. *et al.* Global assessment of trends in wetting and drying over land. *Nature Geosci.* **7**, 716–721 (2014).
10. O’Gorman, P. & Schneider, T. The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Natl Acad. Sci. USA* **106**, 14773–14777 (2009).
11. Orlowsky, B. & Seneviratne, S. I. Elusive drought: Uncertainty in observed trends and short- and long-term CMIP5 projections. *Hydrol. Earth Syst. Sci.* **17**, 1765–1781 (2013).
12. Xoplaki, E., González-Rouco, J. F., Luterbacher, J. & Wanner, H. Wet season Mediterranean precipitation variability: influence of large-scale dynamics and predictability. *Clim. Dyn.* **23**, 63–78 (2004).
13. Steinman, B. *et al.* Ocean–atmosphere forcing of centennial hydroclimate variability in the Pacific Northwest. *Geophys. Res. Lett.* **41**, 2553–2560 (2014).
14. Chen, J. *et al.* Hydroclimatic changes in China and surroundings during the Medieval Climate Anomaly and Little Ice Age: spatial patterns and possible mechanisms. *Quat. Sci. Rev.* **107**, 98–111 (2015).
15. Ljungqvist, F. C., Krusic, P. J., Brattström, G. & Sundqvist, H. S. Northern Hemisphere temperature patterns in the last 12 centuries. *Clim. Past* **8**, 227–249 (2012).
16. Braconnot, P. *et al.* Evaluation of climate models using palaeoclimatic data. *Nature Clim. Change* **2**, 417–424 (2012).
17. Masson-Delmotte, V. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 383–464 (Cambridge Univ. Press, 2013).
18. Trenberth, K. Changes in precipitation with climate change. *Clim. Res.* **47**, 123–138 (2011).
19. Dai, A. Increasing drought under global warming in observations and models. *Nature Clim. Change* **3**, 52–58 (2012).
20. Cook, E. R., Woodhouse, C. A., Eakin, C. M., Meko, D. M. & Stahle, D. W. Long term aridity changes in the western United States. *Science* **306**, 1015–1018 (2004).
21. Cook, E. R. *et al.* Asian monsoon failure and megadrought during the last millennium. *Science* **328**, 486–489 (2010).
22. Evans, M. N., Smerdon, J. E., Kaplan, A., Tolwinski-Ward, S. E. & González-Rouco, J. F. Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors. *Geophys. Res. Lett.* **41**, 9127–9134 (2014).
23. Schmidt, G. A. *et al.* Climate forcing reconstructions for use in PMIP simulation of the last millennium (v1.1). *Geosci. Model Dev.* **5**, 185–191 (2012).
24. Broecker, W. S. & Putnam, A. E. Hydrologic impacts of past shifts of Earth’s thermal equator offer insight into those to be produced by fossil fuel CO₂. *Proc. Natl Acad. Sci. USA* **110**, 16710–16715 (2013).
25. Schneider, T., Bischoff, T. & Haug, G. H. Migrations and dynamics of the intertropical convergence zone. *Nature* **513**, 45–53 (2014).
26. DeAngelis, A. M., Qu, X., Zelinka, M. D. & Hall, A. An observational radiative constraint on hydrologic cycle intensification. *Nature* **528**, 249–253 (2015).
27. Mauritsen, T. *et al.* Climate feedback efficiency and synergy. *Clim. Dyn.* **41**, 2539–2554 (2013).
28. Zhang, X. *et al.* Detection of human influence on twentieth-century precipitation trends. *Nature* **448**, 461–465 (2007).
29. Allan, R. P. & Soden, B. J. Atmospheric warming and the amplification of precipitation extremes. *Science* **321**, 1481–1484 (2008).
30. Polson, D., Hegerl, G., Allan, R. & Sarojini, B. B. Have greenhouse gases intensified the contrast between wet and dry regions? *Geophys. Res. Lett.* **40**, 4783–4787 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements Funding for this work was provided in part by the Swedish Research Council (grant number C0592401), and the Navarino Environmental Observatory (NEO) (project number 1946322). E.Z.’s contribution is part of the German Cluster of Excellence CLISAP (grant number EXC177). The publication cost was covered by the Bolin Centre for Climate Research, Stockholm University, and the Department of Physical Geography, Stockholm University. This is a contribution to the Past Global Changes (PAGES) 2k Network. We thank U. Büntgen at the Swiss Federal Research Institute WSL, and H. Grudd at the Swedish Polar Research Secretariat, for comments on the manuscript.

Author Contributions F.C.L. and P.J.K. designed the study from an original idea by F.C.L. and P.J.K., with input from H.S.S., E.Z., G.B. and D.F. F.C.L. and P.J.K. collected all the proxy data and H.S.S. screened the records for dating uncertainties. P.J.K. produced the software used for the analyses with input from the co-authors. E.Z. provided the model data and calculated the correlation decay length information. All authors contributed to discussion and interpretation of the results. F.C.L., P.J.K. and D.F. wrote the paper with input from the other co-authors.

Author Information Data and code is digitally archived at the NOAA Paleoclimatology/World Data Center for Paleoclimatology (<https://www.ncdc.noaa.gov/paleo/study/19725>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.C.L. (fredrik.c.l@historia.su.se).

METHODS

The proxy data collection. We systematically searched the scientific literature for proxy records of hydroclimate and temperature. We retrieved all records fulfilling the following criteria:

- (1) Temporal coverage. All proxy records must cover at least the period 1000–1899.
- (2) Temporal resolution. All proxies must have at least, on average, two data points per century between 1000 and 1899 and no more than two centuries over the 1000–1899 period may have fewer. In the data subset containing only hydroclimate records resolved decadal or better, we required that all proxies have, on average, one data point per century between 1000 and 1899.
- (3) Publication requirements. We use only proxy records that have been published in the scientific literature.
- (4) Requirements regarding the climate signal. The proxies must, according to their published description, contain a hydrological signal described as precipitation, drought, moisture balance, stream flow, lake level change, bog surface wetness, flooding or similar hydrological information.
- (5) Geographical requirements. All the proxies must be located in, or nearly in, the Northern Hemisphere. Proxies just south of the Equator (by a few degrees) may be included if they contribute to gridded values north of the Equator.

The majority of records were obtained from public repositories (for example, <http://www.ncdc.noaa.gov/paleo/> and <http://www.pangaea.de/>). Those records not in the public domain were acquired either by direct request from their authors, or digitized from figures in the original articles (Supplementary Tables 1 and 2). The proxy data were divided into seven categories: (1) ice cores, (2) peat cores, (3) tree rings, (4) speleothems, (5) documentary, (6) lake sediments, and (7) marine sediments. In cases where there exist several versions of a proxy record from the same site, preference was given to the latest published version. The collected proxy records were divided into three simplified categories of seasonal response: annual, winter and summer, with spring and early autumn considered as part of the summer season. In cases when no clear information about a proxy's seasonality could be found, either in the original article or in subsequent publications, we assumed the seasonality to be annual.

Geochronological screening. We only included proxy records with age models constrained by at least one dating point after 1900 (D1), another between 1000 and 1900 (D2) and one dating point before 1000 (D3). The number of years between D1 and D2 and D2 and D3 should be less than 1,000 years.

The dating criteria are:

- (1) The top of the core exists (that is, year of collection) or, if it does not, there must be a dating point (for example, ^{137}Cs , ^{210}Pb) to verify the age (A1 is the top age).
- (2) There should be at least one dating point (^{14}C , U/Th, tephra, ^{210}Pb) between 800 and 1999 (A2).
- (3) There should be at least one dating point between 800 and 1 (A3).
- (4) The spacing between these three dating points should be less than 800 years ($A1 - A2 < 800$ years, $A2 - A3 < 800$ years).
- (5) A1, A2 and A3 cannot be bulk datings.
- (6) The acceptable dating errors for A1, A2 and A3 must be < 200 years (2σ).

For the decadal- or better-resolved subset, each record must have at least one dating point per full century.

Data treatment. Prior to computing centennial anomalies, all proxy records with irregularly spaced time steps were converted into time series with annually spaced time steps using simple linear interpolation, then smoothed by a cubic smoothing spline having a 50% frequency response at 100 years. The smoothed proxy series are transformed to standard normal deviates with respect to the 1000–1899 baseline period. From the resulting centennial anomalies every 25th value is extracted from each century with at least 85 (interpolated) values. From those proxy records spanning the complete 800–1999 period, we obtain 45 centennial anomalies. Of these 45 centennial anomalies the 12 representing the hundred-year periods 800–899, 900–999, 1000–1099, 1100–1199, ..., 1900–1999, are those presented in the majority of figures throughout this Letter (Extended Data Fig. 8b). However, all 45 centennial anomalies are used in the correlation experiments as well as in Fig. 2b and c. For the subset including only hydroclimate records decadal or better, we obtained decadal anomalies using the same method as described above but instead using a spline with a 50% frequency response at 10 years.

Correlation decay length. Because individual proxy records can contain noise, a more robust signal is presumably obtained by producing an average of multiple proxies from within a geographic area where hydroclimate can be assumed to be coherent. Unfortunately, the size of such geographic areas, on centennial timescales, is not well constrained owing to the limited length and distribution of instrumental observations. This deficiency required the development of a measure of correlation decay length (CDL)³¹ of hydroclimate^{32,33} on centennial timescales. Encouraged by the agreement found between instrumental precipitation data on annual and decadal timescales³⁴ and simulated values from the ECHO-G model³⁵

on the same timescales, we used the simulated centennial values from the ECHO-G model to calculate a conservative estimate of the spatial auto-correlation function for hydroclimate on centennial timescales by the following equation:

$$C(j, L) = \frac{1}{N} \sum_i \left\{ \frac{1}{M} \sum_{\substack{t, k, l \\ L - \Delta L \leq d \leq L + \Delta L}} f(t, i, j) f(t, k, l) \right\} \quad (1)$$

where t is the time index, i and j are the longitude index and the latitude index of the grid, respectively; $f(t, i, j)$ are the deviations from the local long-term mean at time t and at a grid-cell (i, j) divided by the local temporal standard deviation (that is, standardized to one standard deviation), d is the spatial separation between grid cells (i, j) and (k, l) . The sum of the product $f(t, i, j)f(t, k, l)$ over time and space is restricted to the grid cells (k, l) that are separated from the grid cell (i, j) by the distance between $L - \Delta L$ and $L + \Delta L$, where ΔL is 100 km. This sum is divided by the number of occurrences of grid-cell pairs. The outer sum is calculated over all grid cells located on an index circle j , and then this sum is divided by that number of grid cells. It is thus assumed that the spatial autocorrelation is isotropic and zonally homogeneous, that is, it depends only on length L and on latitude j . The assumption of isotropy and zonal homogeneity is an assumption that is not completely fulfilled for all terrestrial locations. However, this approach yields an approximate and average estimation of the decorrelation length at multi-decadal to centennial timescales, and this estimation amounts to a few hundred kilometres (Extended Data Fig. 1a). The precise value of the decorrelation length does not affect the basic features of the reconstruction. A simplified CDL function of hydroclimate variability on centennial timescales was derived from the output of the climate model ECHO-G³⁵ over the period 1000–1990 to calculate the maximum search distances as a function of latitude in this study (Extended Data Fig. 1b).

Averaging, weighting and gridding. We use the distance function described above, derived from the estimated centennial hydroclimate CDL, to define the search radius for averaging proxy anomalies (Extended Data Fig. 1a and b). Finding the decorrelation of hydroclimate increases with distance from a search centre, we assign a weight for each proxy record, for any given latitude, decreasing from 1 at the centre to $e^{-2} \approx 0.14$ at the search periphery using the following Gaussian weight function:

$$w = e^{-2x^2/R^2} \quad (2)$$

Here, w is the weight assigned to a proxy value located at x distance from a grid node, and R is the search radius defined by equation (1). Every reconstructed grid cell location is required to have three or more proxy centennial anomalies within the search radius. This requirement effectively reduces the total number of proxies used in the experiment, although some regions remain densely covered (such as North America, China and Europe; Extended Data Fig. 8a). Weighted means are computed for all 45 centennial anomalies and are those used in the correlation analyses. All gridding and weighting of proxy (and model) data are performed using the method described above for producing the proxy local weighted averages, although the search area centres are moved to each intersection of a $1^\circ \times 1^\circ$ longitude by latitude grid superimposed over a polar projection of the Northern Hemisphere. Maps of centennial gridded proxy hydroclimate variability are shown in Fig. 2a, and time series corresponding to the fraction of land area at or above a given wetness or dryness threshold are shown in Fig. 2b. The weighted and gridded proxy temperature reconstruction shown in Extended Data Fig. 6a and used in Fig. 3 is produced in the same manner as the hydroclimate reconstructions, with the exception of the CDL function used for weighting and gridding. Detailed descriptions of the CDL, the proxy weighting function, and the search distance function used for reconstructing Northern Hemisphere centennial temperature variations are given in ref. 15.

Climate model data. We have used simulations from six coupled atmosphere-ocean general circulation model simulations that are part of the Climate Model Intercomparison project Phase 5 (CMIP5)³⁶. The simulations denoted as 'past1000' cover the period 850–1850, whereas the 'historical' simulations cover the period 1851–2005. The 'past1000' simulations are sometimes performed using a lower resolution or simplified versions of the more comprehensive models used for the 'historical' simulations. We have restricted our analysis to those six simulations—bcc-csm1-1³⁷, CCSM4³⁸, IPSL-CM5A-LR³⁹, GISS-E2R⁴⁰, HadCM3⁴¹, and MPI-ESM-P⁴²—conducted with the same model versions for both periods and stitch the 'past1000' and 'historical' simulations that carry the same ensemble run-initialisation-physics (rip) identifier in the CMIP5 repository (Table 1). However, the simulation conducted with the MIROC-ESM model was not used owing to a documented, unrealistic, long-term positive temperature trend in global mean annual near-surface temperature over the past millennium⁴³.

The models were driven by estimations of external forcing, including variations in the orbital changes, total solar irradiance, volcanism, atmospheric trace

gases and land-use changes. The CMIP5 Consortium recommended similar, but somewhat different, options for these external forcings^{23,44} and the modelling groups did not always use the same options for all CMIP5 simulations. From each of the six simulations we extracted the annual mean precipitation and temperature at the model grid-cell location nearest to our proxy locations. Once the selection was made, an equal length extraction was performed to make all comparisons between model and proxy as equitable as possible. The smoothing, weighted averaging and gridding of all model data followed the same protocol as described above for the proxy data. The centennial forcing anomalies in Fig. 3 are produced from time series of solar⁴⁵, volcanic⁴⁶ and CO₂ forcing⁴⁷ the same or similar to those used in the models.

Calculation of correlations. All temporal correlations between two weighted and gridded data sets are calculated by Pearson product-moment correlation coefficient⁴⁸. Assessment of the statistical significance of these correlations is complicated by the fact that any two data sets are both temporally and spatially correlated. We have taken temporal correlation, that is, autocorrelation, into account by using the method of blocked bootstrap (3,001 repetitions)⁴⁹ to provide a more realistic estimate of the correlation's statistical significance. In Extended Data Fig. 4a we show the correlations between the gridded hydroclimate proxy data and corresponding gridded hydroclimate model data and Extended Data Fig. 4b shows the statistical significance of the correlations. In Extended Data Fig. 4c we show the correlations between the gridded hydroclimate proxy data and gridded temperature proxy data and Extended Data Fig. 4d shows the statistical significance of the correlations.

We produce histograms of all cross-correlations in this study (Extended Data Fig. 9a–d) and their Fisher *z* transform⁵⁰ (Extended Data Fig. 9e–h) to assess the normality of their distributions. A slightly bimodal distribution is seen, for both the original and Fisher *z*-transform correlations, between the gridded hydroclimate proxy data and gridded temperature proxy data. This is not what we would expect to see if hydrological and temperature anomalies were unrelated: the distribution of the Fisher *z* transform of the correlations should then be approximately normal and, in particular, unimodal. Instead, the histograms suggest that the distributions of the correlations between proxy hydroclimate and proxy temperature values are a mixture of two unimodal distributions, corresponding to two distinct mechanisms relating hydroclimate and temperature.

Validation tests. To assess the robustness of our gridded weighted hydroclimate reconstruction we have produced a number of 'subset' reconstructions, using different data screening schemes, as well as comparing our multi-proxy reconstruction to other hydroclimate products. We created a subset of proxies with only those records that have decadal to annual resolution and at least one dating point per century. The gridded reconstruction obtained from this subset has less spatial coverage than the reconstruction from the full data set. The median Pearson product-moment correlation coefficient, at grid point level, between the two reconstructions (using 45 centennial values, lagged by 25 years) is $r = 0.97$ ($p < 0.01$), suggesting that any potential dating uncertainties that exist in the full data set have a small influence on our overall results (Extended Data Table 1). Furthermore, we correlated the full reconstruction with reconstructions produced excluding one of the seven proxy types at the time, which notably also involves excluding the numerous but less well age-constrained lake sediment records (Extended Data Table 1).

Moreover, we produced similar tests excluding/retaining proxy records reflecting quantitative estimates of precipitation as well as excluding/retaining proxy records with annual signal, summer signal, and winter signal, respectively. Finally, we correlated our reconstruction from the full data set as well as the reconstruction from the more strictly screened data set resolved decadal or better, set against the North American Drought Atlas²⁰ and the Monsoon Asia Drought Atlas²¹. At grid point level the median Pearson's correlation is $r = 0.34$ ($p < 0.15$) and $r = 0.50$ ($p < 0.03$), respectively. These significant, but not very high, correlation values are acceptable considering our hydroclimate reconstruction metric is not an equivalent measure of the PDSI^{51,52} and furthermore that the PDSI reconstructions are for the summer season only.

Potential spectral biases in proxy data. The Earth's climate system contains variability from sub-daily to Milankovitch to deep geological timescales^{17,53–55}. It has been known for a long time that proxy records may record climatic variation occurring on different wavelengths (that is, interannual to decadal to centennial to millennial) with varying fidelity resulting in disproportionate manifestations of the variance across these different wavelengths^{56–62}. Such 'spectral biases' are a type of noise that, if not considered or corrected, potentially challenge a precise quantitative assessment of the continuum of climate variability. Using instrumental records to characterize and subsequently correct the possible spectral biases in proxy records should yield more accurate reconstructions of climate. However, we note that this endeavour to characterize proxy biases is questionable, because the true long-term characteristics of the climate system remain broadly unknown, which is part of the motivation to develop high-quality proxy data and reconstructions.

Knowledge of proxy data sets is, however, sufficiently advanced to state that individual proxy types may be characterized by one or more biases. For example, tree-ring data, arguably the most intensively investigated proxy archive for the past millennium in both the temporal and frequency domains, are known to have two potentially opposing biases. On the one hand, owing to a 'biological memory', tree-ring width data especially contain somewhat less high-frequency (interannual) variability than instrumental temperature (or precipitation) data^{59–62}. On the other hand, the necessary removal of the long-term age/size-related trends in tree-ring records, if performed using certain techniques, results in the so-called 'segment length curse'⁶³ that limits our ability to preserve trends/variability exceeding the length of the mean length of the individual tree-ring sequences. In this case, diminished low-frequency (centennial to millennial) variability would be expected.

There are, however, different techniques that can be applied to overcome or mitigate such spectral biases. The removal of the age-related trends using a single population estimate as in the Regional Curve Standardisation technique has been shown to break the 'segment length curse' and to preserve long wavelengths better^{63–65}. Similarly, the biological memory can be estimated and removed by autoregressive modelling and the resulting time series can be assigned the spectral properties expected from the instrumental data sets⁵⁶. Analyses of long-term instrumental data sets are limited to the past couple of centuries, which does not, as also mentioned in the main text, allow a full validation of the low-frequency (for example, centennial to millennial timescale) components of proxy data or reconstructions.

In most proxy records, just as with the tree-ring data, the low-frequency variability can be affected by nonlinear processes such as the compression or mixing of layers, a time-varying temperature–precipitation relationship, anthropogenic impacts on the local environment and, in the case of ice-core data, ice-flow movements in the ice cap. It should be noted that the interpolation of non-annual records to annual resolution also results in a spectral bias towards lower frequencies⁶⁶. Given the uncertainties back in time in many individual proxy records, both in the high- and low-frequency signal, a multi-proxy approach, using the average signal from a number of disparate proxy records, is arguably best suited for the study of low-frequency climate variability.

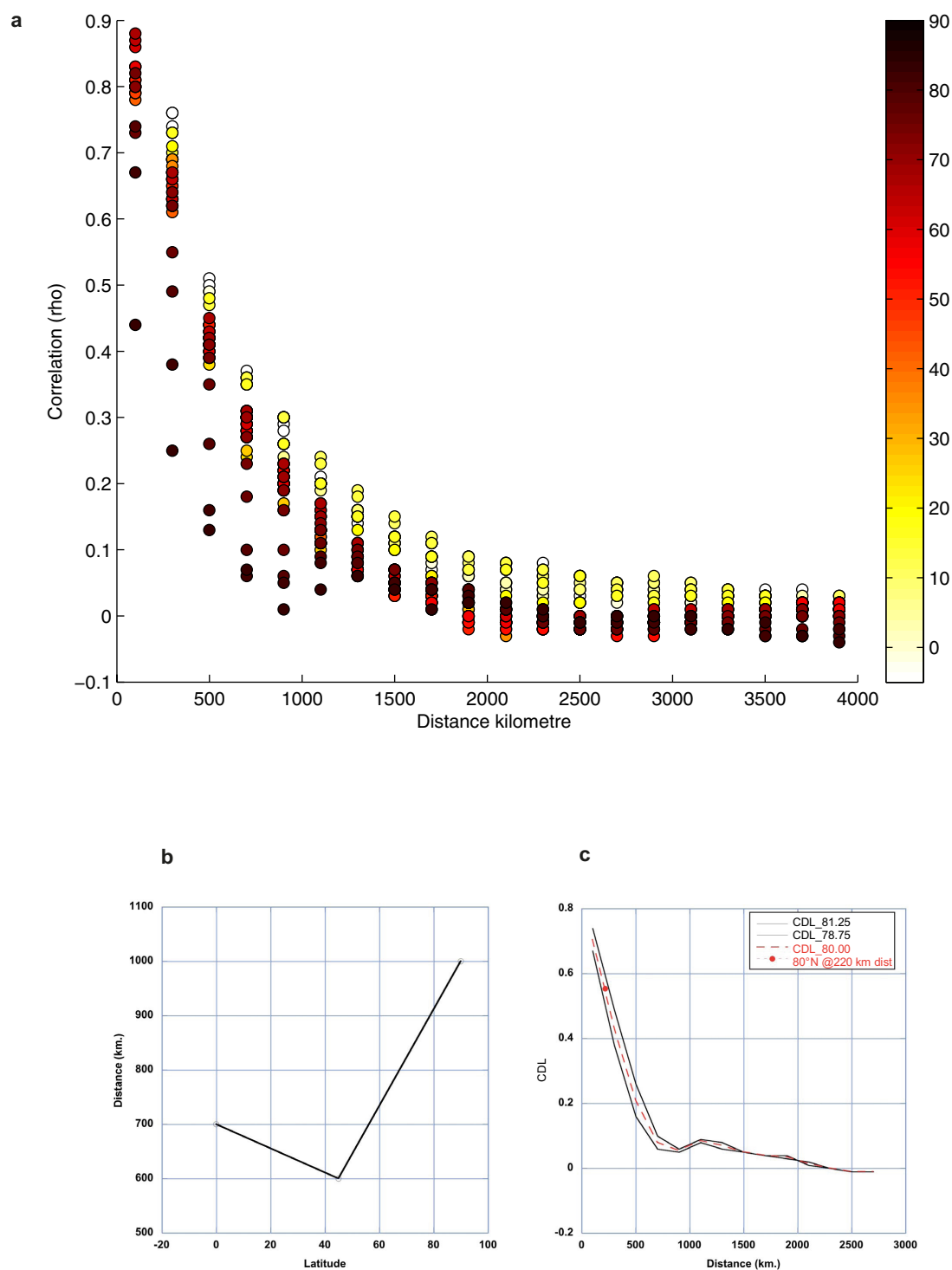
In multi-proxy reconstructions, where the aim is to obtain both the high (annual) and low (centennial) frequency signal, any spectral bias in the proxies, influencing the ratio of the high- and low-frequency signal, is a problem of major importance. Our aim, however, is only to obtain the signal on centennial and longer timescales. By standardizing centennial mean values, proxy records that have both strong and weak low-frequency signals are transformed to records of just centennial-scale information of the same amplitude. This transformation essentially eliminates the influence of spectral biases by making irrelevant whatever prior ratio of high- and low-frequency information had existed in the proxy before standardization. Even if a proxy does not capture the full amplitude of centennial-scale variability (for example, tree-ring width chronologies built by many short individual series and/or limited replication⁶³) the impact is small because the standardization process gives all proxies the same amplitude of centennial-scale variability. As long as the proxies have some centennial-scale variability (which is the only variability we retain for our primary reconstruction) the proxies are useful to us.

Our different sensitivity tests (Extended Data Table 1), removing various proxy types from the complete data set and correlating their resulting reconstruction with the full reconstruction, demonstrate a very limited impact of potential spectral biases from particular proxy types on our full reconstruction. These experiments produce remarkably similar hydroclimate reconstructions, suggesting that the purported negative effect of 'mixing' proxies with potential spectral biases is not large (if not altogether mitigated by the standardization). Especially important in this regard are results from those tests in which all the records with less than decadal resolution, with only centennial dating control, and those in which all the tree-ring records and all the lake sediment records are excluded, respectively. Not surprisingly, the reduction in proxy coverage gave rise to some spatially local differences; however, the overall patterns remained the same. Most of the changes seen are less likely to be attributable to spectral biases in the proxies, but rather to a diminishment in the considered number of proxies, and hence greater noise in the reconstruction. We mitigate the impact of proxy noise in our reconstruction approach by requiring multiple records within the search radius for every grid point reconstructed.

Data and code availability. All the proxy data and programs used to perform the experiments herein have been made publicly available at NOAA Paleoclimatology/World Data Center for Paleoclimatology (<https://www.ncdc.noaa.gov/paleo/study/19725>). The Source Data we used to plot each figure in the Letter are also stored there. The suggested computing environment and dependencies are: Mac OSX 10.6 or greater, Generic Mapping Tools (GMT) 4.5.6 or later⁶⁷, Ghostscript 9.10, and Matlab2007b or later. A detailed description of the software used, with example data and runtime commands is supplied with the archived data.

The climate model data included in the simulations of the Climate Model Intercomparison Project Version 5 (CMIP5) can be downloaded from any of the nodes of the Earth System Grid Federation (registration is required and a password is provided after registration), for example from: <https://esgf-data.dkrz.de/search/cmip5-dkrz/>. The menu on the left side of the page allows the user to select the 'experiment' (in our case, 'past1000' and 'historical'), the model (for example, MPI-ESM-P), the variable (pr = precipitation), and the time frequency (mon = monthly). A click on the search button displays all matching results. The data download can be performed interactively by clicking on the links of the individual files or by downloading a Unix script 'wget', which can be locally run on a Unix computer. All files are written in 'netcdf' format, which also includes the metadata information, and that can be accessed with a variety of software tools, for instance with the public statistical packages R, Climate Data Operators (CDO) or netcdf Operators (NCO). Data of the ECHO-G simulation can be retrieved from the CERA data bank of the German Climate Computing Centre (DKRZ) at <http://cera-www.dkrz.de/>. An account on the CERA data bank is required to access the data and can be obtained by sending an email to data@dkrz.de. CERA will then provide a personal account and password as soon as possible. Once an account has been created, the data of the ECHO-G simulation can be accessed.

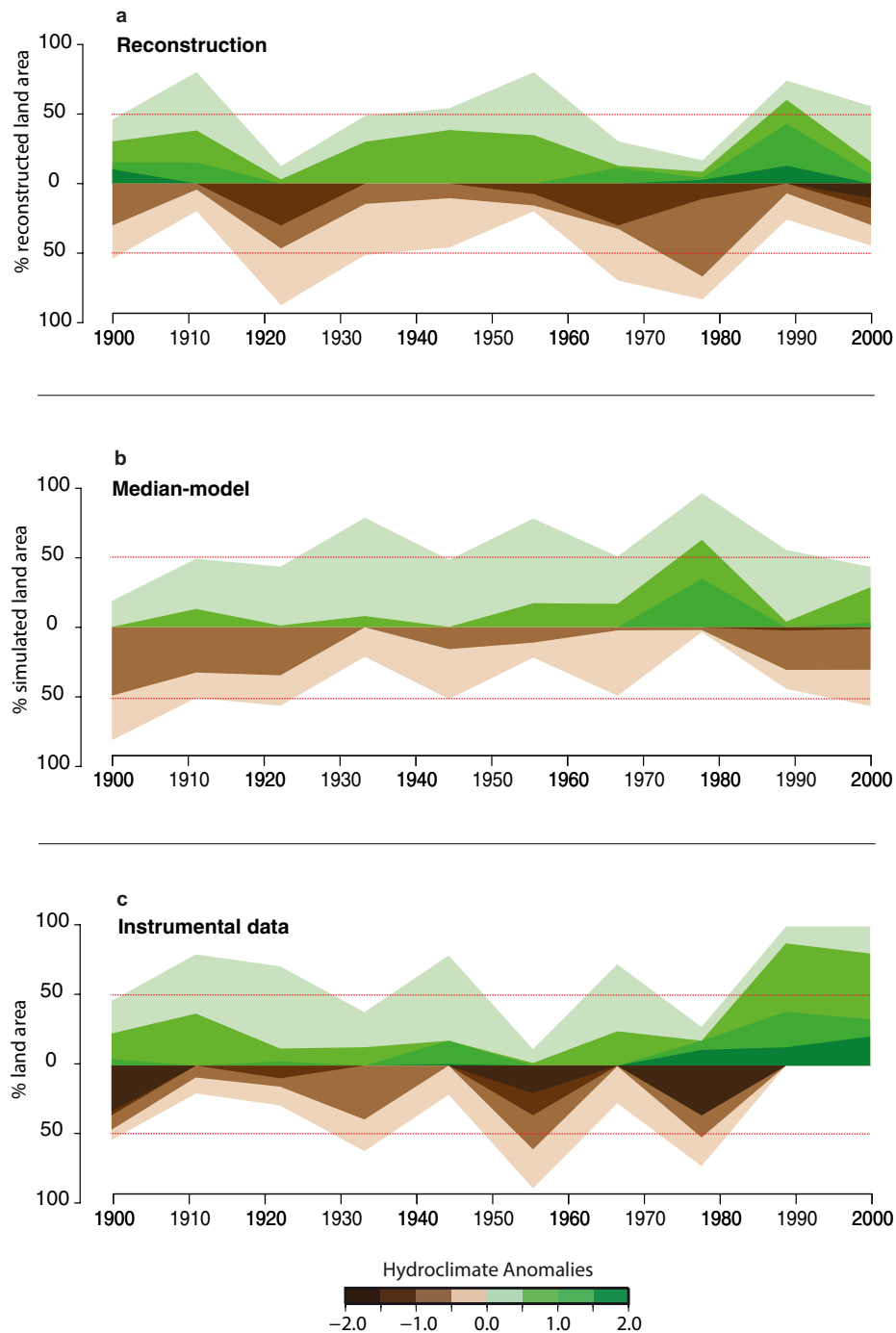
31. Jones, P., Osborn, T. & Briffa, K. Estimating sampling errors in large-scale temperature averages. *J. Clim.* **10**, 2548–2568 (1997).
32. Datta, S., Jones, W. L., Roy, B. & Tokay, A. Spatial variability of surface rainfall as observed from TRMM field campaign data. *J. Appl. Meteorol.* **42**, 598–610 (2003).
33. Wan, H., Zhang, X., Zwiers, F. W. & Shiogama, H. Effect of data coverage on the estimation of mean and variability of precipitation at global and regional scales. *J. Geophys. Res.* **118**, 534–546 (2013).
34. Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Clim.* **25**, 693–712 (2005).
35. González-Rouco, F., Beltrami, H., Zorita, E. & von Storch, H. Simulation and inversion of borehole temperature profiles in surrogate climates: spatial distribution and surface coupling. *Geophys. Res. Lett.* **33**, L01703 (2006).
36. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
37. Xin, X., Wu, T. & Zhang, J. Introduction of CMIP5 simulations carried out with the climate system models of Beijing Climate Center. *Adv. Clim. Change Res.* [in Chinese] **4**, 41–49 (2013).
38. Landrum, L. *et al.* Last millennium climate and its variability in CCSM4. *J. Clim.* **26**, 1085–1111 (2013).
39. Schmidt, G. A. *et al.* Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.* **6**, 141–184 (2014).
40. Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F. B. & Phipps, S. J. Separating forced from chaotic climate variability over the past millennium. *J. Clim.* **26**, 6954–6973 (2013).
41. Dufresne, J.-L. *et al.* Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim. Dyn.* **40**, 2123–2165 (2013).
42. Giorgetta, M. A. *et al.* Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J. Adv. Model. Earth Syst.* **5**, 572–597 (2013).
43. Sueyoshi, T. *et al.* Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM. *Geosci. Model Dev.* **6**, 819–836 (2013).
44. Schmidt, G. A. *et al.* Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geosci. Model Dev.* **4**, 33–45 (2011).
45. Steinhilber, F., Beer, J. & Frohlich, C. Total solar irradiance during the Holocene. *Geophys. Res. Lett.* **36**, L19704 (2009).
46. Gao, C., Robock, A. & Ammann, C. Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. *J. Geophys. Res.* **113**, D23111 (2008).
47. Petit, J.-R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
48. Pearson, K. Notes on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **58**, 240–242 (1895).
49. Wilks, D. S. Resampling hypothesis tests for autocorrelated fields. *J. Clim.* **10**, 65–82 (1997).
50. Fisher, R. A. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10**, 507–521 (1915).
51. Palmer, W. C. *Meteorological Drought* (US Department of Commerce Research Paper 45, 1965).
52. Dai, A. G., Trenberth, K. E. & Qian, T. T. A global dataset of Palmer drought severity index for 1870–2002: relationship with soil moisture and effects of surface warming. *J. Hydrometeorol.* **5**, 1117–1130 (2004).
53. Lamb, H. H. *Climate: Present, Past and Future* Vols 1 and 2 (Methuen, 1972–1977).
54. Bradley, R. S. *Paleoclimatology: Reconstructing Climates of the Quaternary* (Academic, 1999).
55. Huybers, P. & Curry, W. Links between annual, Milankovitch and continuum temperature variability. *Nature* **441**, 329–332 (2006).
56. Meko, D. M. *Applications of Box-Jenkins methods of time-series analysis to the reconstruction of drought from tree-rings*. PhD dissertation, Univ. Arizona (1981).
57. Guiot, J. The extrapolation of recent climatological series with spectral canonical regression. *J. Climatol.* **5**, 325–335 (1985).
58. Osborn, T. J. & Briffa, K. R. Revisiting timescale dependent reconstruction of climate from tree-ring chronologies. *Dendrochronologia* **18**, 9–25 (2000).
59. Franke, J., Frank, D., Raible, C., Esper, J. & Brönnimann, S. Spectral biases in tree-ring climate proxies. *Nature Clim. Change* **3**, 360–364 (2013).
60. Büntgen, U. *et al.* Tree-ring amplification of the early nineteenth-century summer cooling in central Europe. *J. Clim.* **28**, 5272–5288 (2015).
61. Esper, J. *et al.* Signals and memory in tree-ring width and density data. *Dendrochronologia* **35**, 62–70 (2015).
62. Zhang, H. *et al.* Modified climate with long term memory in tree ring proxies. *Environ. Res. Lett.* **10**, 084020 (2015).
63. Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A. & Funkhouser, G. The 'segment length curse' in long tree-ring chronology development for palaeoclimatic studies. *Holocene* **5**, 229–237 (1995).
64. Esper, J., Cook, E. R. & Schweingruber, F. H. Low-frequency signals in long tree-ring chronologies and the reconstruction of past temperature variability. *Science* **295**, 2250–2253 (2002).
65. Esper, J., Cook, E. R., Krusic, P. J., Peters, K. & Schweingruber, F. H. Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies. *Tree-ring Res.* **59**, 81–98 (2003).
66. Schulz, M. & Stattegger, K. SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series. *Comput. Geosci.* **23**, 929–945 (1997).
67. Wessel, P. & Smith, W. H. F. New, improved version of the Generic Mapping Tools released. *Eos* **79**, 579 (1998).



Extended Data Figure 1 | Estimated correlation decay length values.

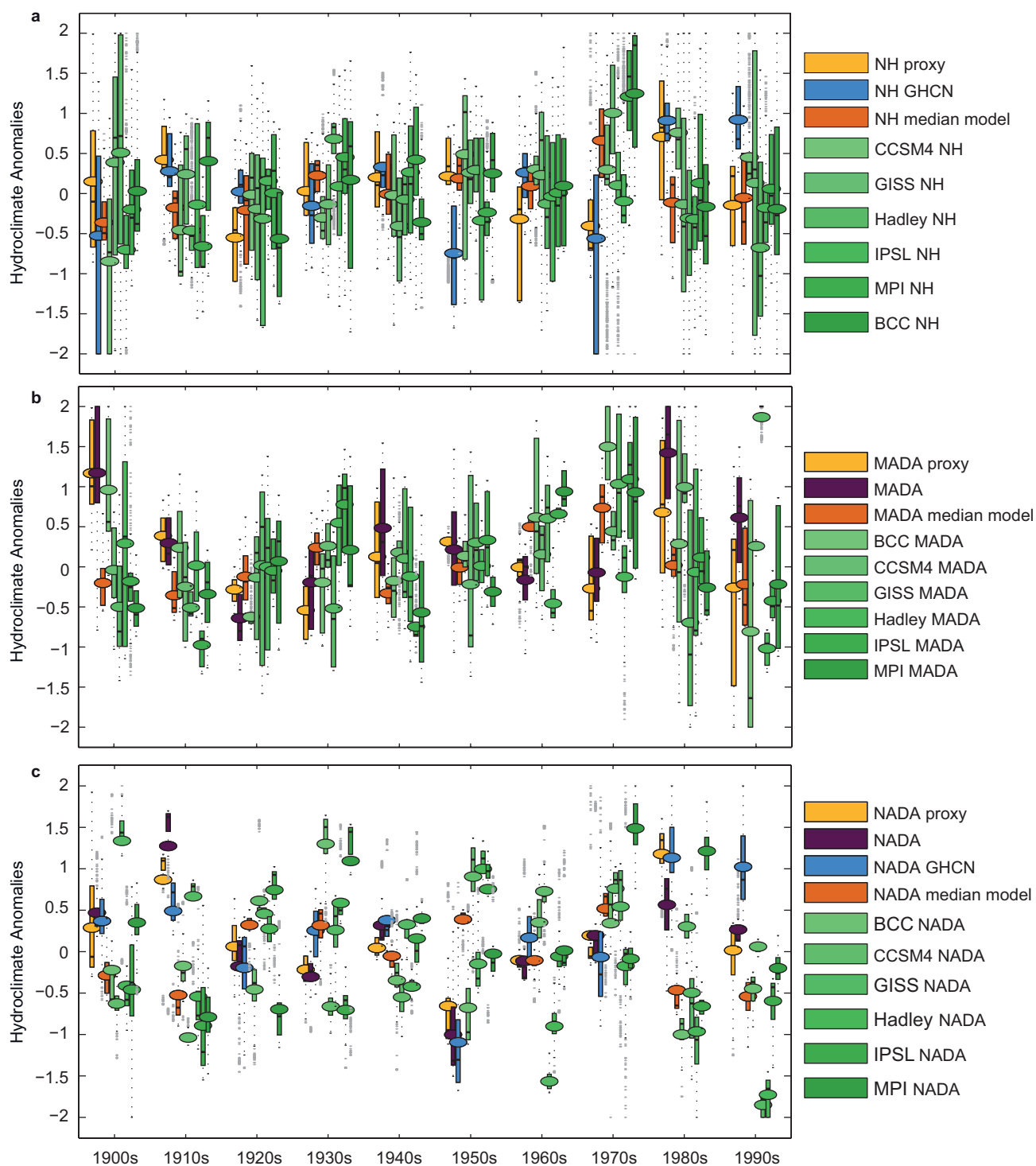
a, The spatial decorrelation function ρ for centennial mean precipitation estimated from the output of the climate model ECHO-G³⁵ over the period 1000–1990, following the procedure described in the Methods for the estimation of the centennial correlation decay length for hydroclimate variability. Distance is the correlation decay length from one point in kilometres. The different colours represent the latitudinal bands.

b, An example of the estimated spatial autocorrelation function for centennial mean values of precipitation at latitudes 81.25° N, 80.00° N and 78.75° N, respectively, with the decorrelation length for latitude 80° N indicated in red. **c**, The simplified CDL function of hydroclimate variability at centennial timescales, derived from panel **a**, is used throughout the study to calculate the maximum search distances as a function of latitude.



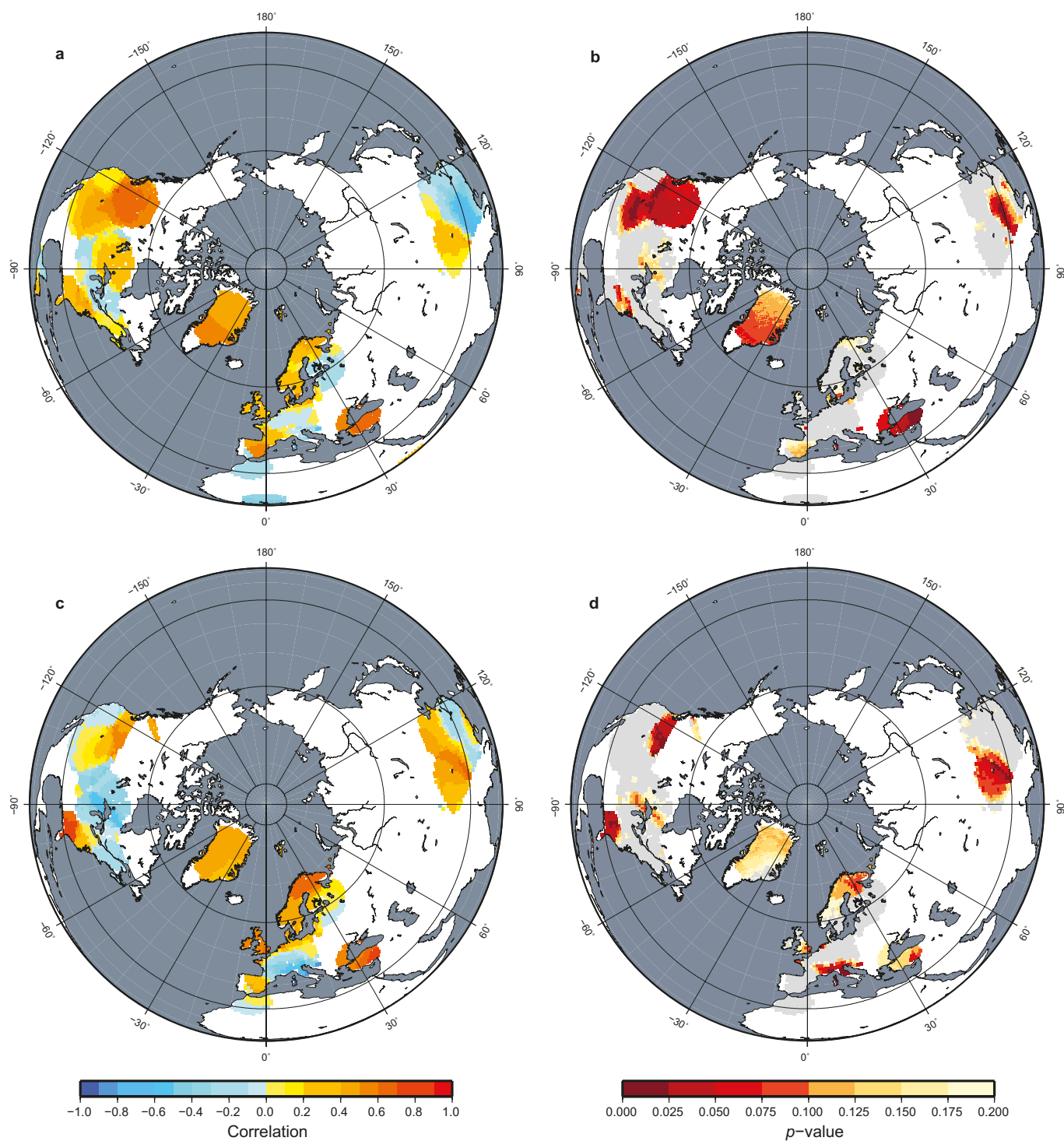
Extended Data Figure 2 | The fraction of land area, expressed as decadal means for 1900–1999, exceeding a given wetness or dryness threshold in the gridded reconstruction, model simulations, and instrumental precipitation data. a, Weighted gridded proxy reconstruction derived from the subset containing only hydroclimate records resolved decadal or better. **b,** The same as **a** but for median-model anomaly values of

annual precipitation. **c,** The same as **a** and **b** but for the Global Historical Climatology Network $5^\circ \times 5^\circ$ (GHCN5) instrumental annual precipitation data. All decadal average values are standardized over the 1910–1979 period, and model and instrumental values are extracted from grid cells covered by gridded reconstructed values. The red horizontal bars denote the 50% levels.



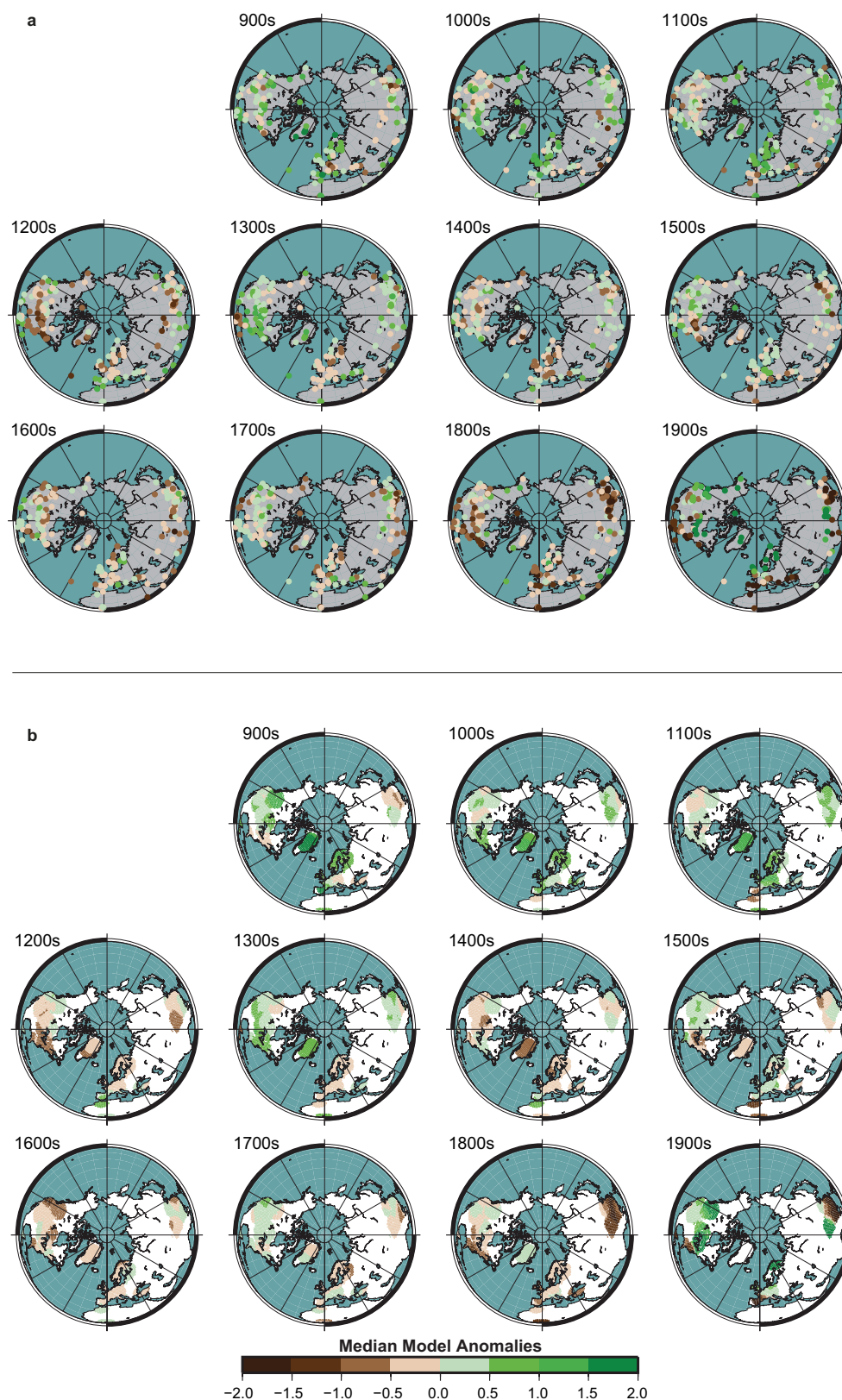
Extended Data Figure 3 | Boxplots showing decadal anomaly values of instrumental data, our gridded reconstruction, drought atlas data, and model simulation of precipitation over the 1900s. a, Comparison of decadal anomalies between our gridded Northern Hemisphere (NH) hydroclimate reconstruction, the gridded $5 \times 5^\circ$ GHCN5 instrumental annual precipitation anomalies, the six individual model simulations (see Table 1) of annual precipitation and their median. **b,** Comparison of decadal anomalies between the Monsoon Asia Drought Atlas (MADA)²¹ and the corresponding domain in our gridded Northern Hemisphere hydroclimate reconstruction, the GHCN5 instrumental

annual precipitation data set, and in the six individual model simulations of annual precipitation and their median. **c,** Comparison of decadal anomalies between the North American Drought Atlas (NADA)²⁰ and the corresponding domain in our gridded Northern Hemisphere hydroclimate reconstruction, the GHCN5 instrumental annual precipitation data set, and in the six individual model simulations of annual precipitation and their median. The oval circles represent the mean, the small blank horizontal bar represents the median, the length of the bars represents the quartile range, and the dark grey dots represent the two standard deviation intervals, whereas the light grey dots represent outliers.



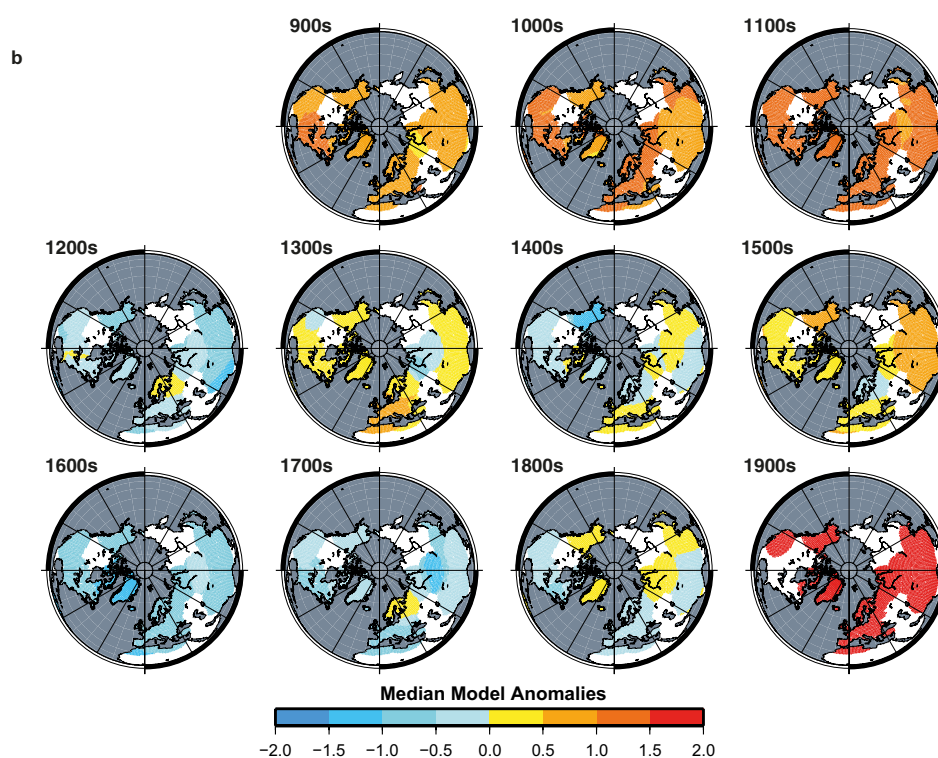
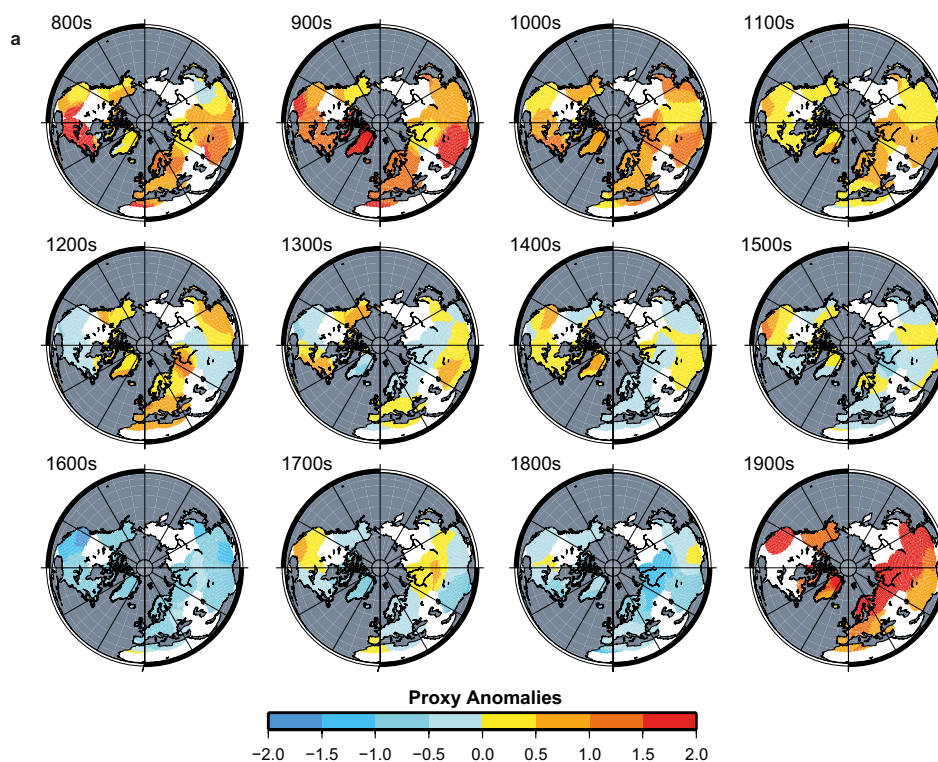
Extended Data Figure 4 | Correlations between gridded proxy and model hydroclimate anomalies, and gridded hydroclimate temperature proxy anomalies. **a**, Correlations between 45 centennial, lagged 25 years, weighted gridded proxy hydroclimate anomalies and their corresponding median total annual precipitation anomalies from six CMIP5 models, listed in Table 1, over the past twelve centuries. **b**, The

Z-transformed block bootstrap p -values of the correlations shown in panel **a**. **c**, Correlations between 45 centennial, lagged 25 years, weighted gridded proxy hydroclimate anomalies and weighted gridded proxy temperature anomalies. **d**, The Z-transformed block bootstrap p -values of the correlations shown in panel **c**. Areas shown in grey in **b** and **d** have insignificant correlations.



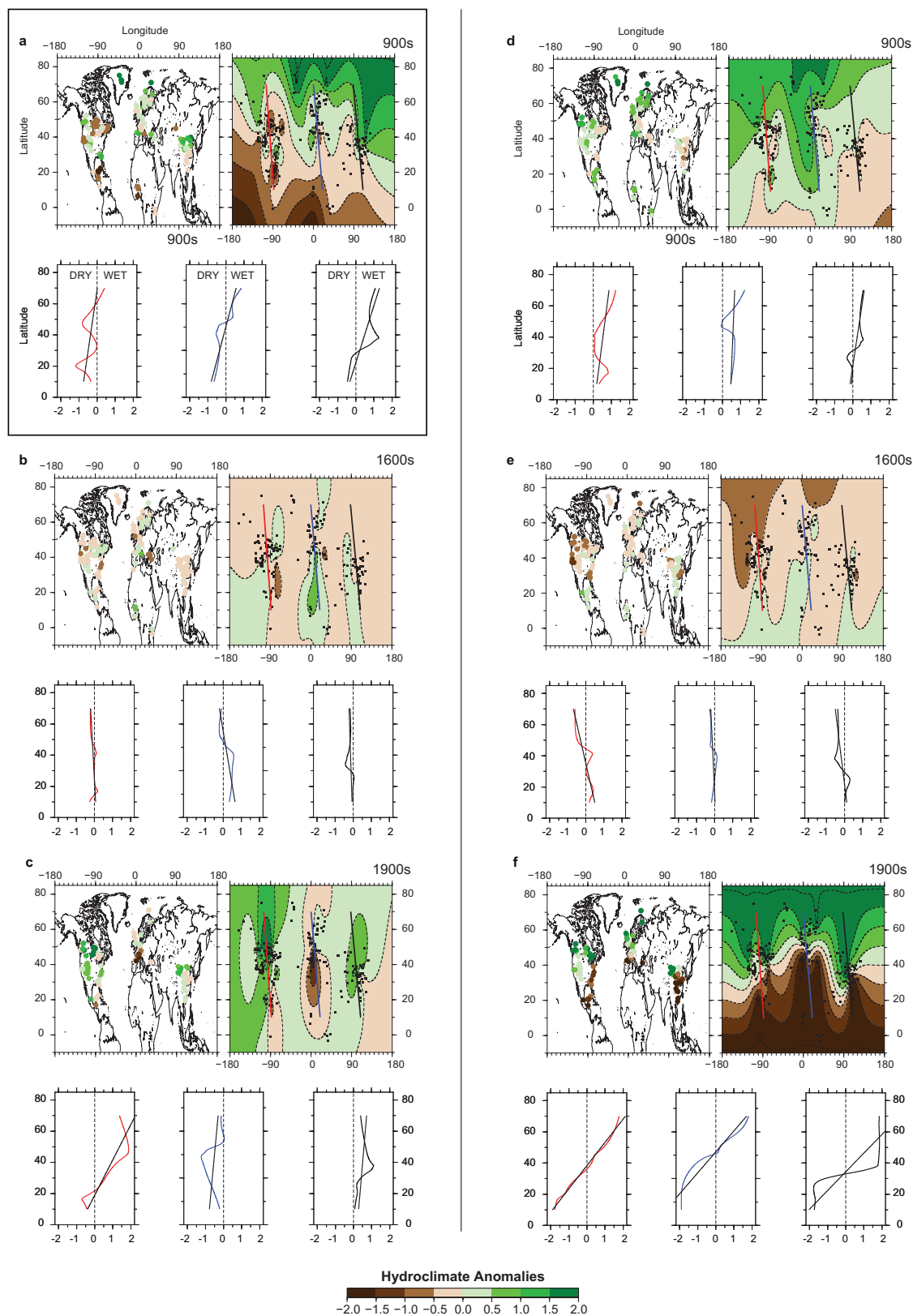
Extended Data Figure 5 | Simulated median values of annual precipitation from six atmosphere–ocean coupled general circulation models. a, Raw, centennial, model anomaly median values calculated and treated and plotted in the same way as the hydrological proxy data. Only values from the same grid cells that are covered by proxy records are extracted (Methods). Anomalies are shown relative to the centennial

mean and standard deviation over the eleventh–nineteenth centuries. The colour scale in both panels is truncated at -2 and 2 . **b** Gridded, weighted, values for the same data over land areas with at least three independent grid values within the estimated centennial correlation decay length for centennial-scale hydrological variability.



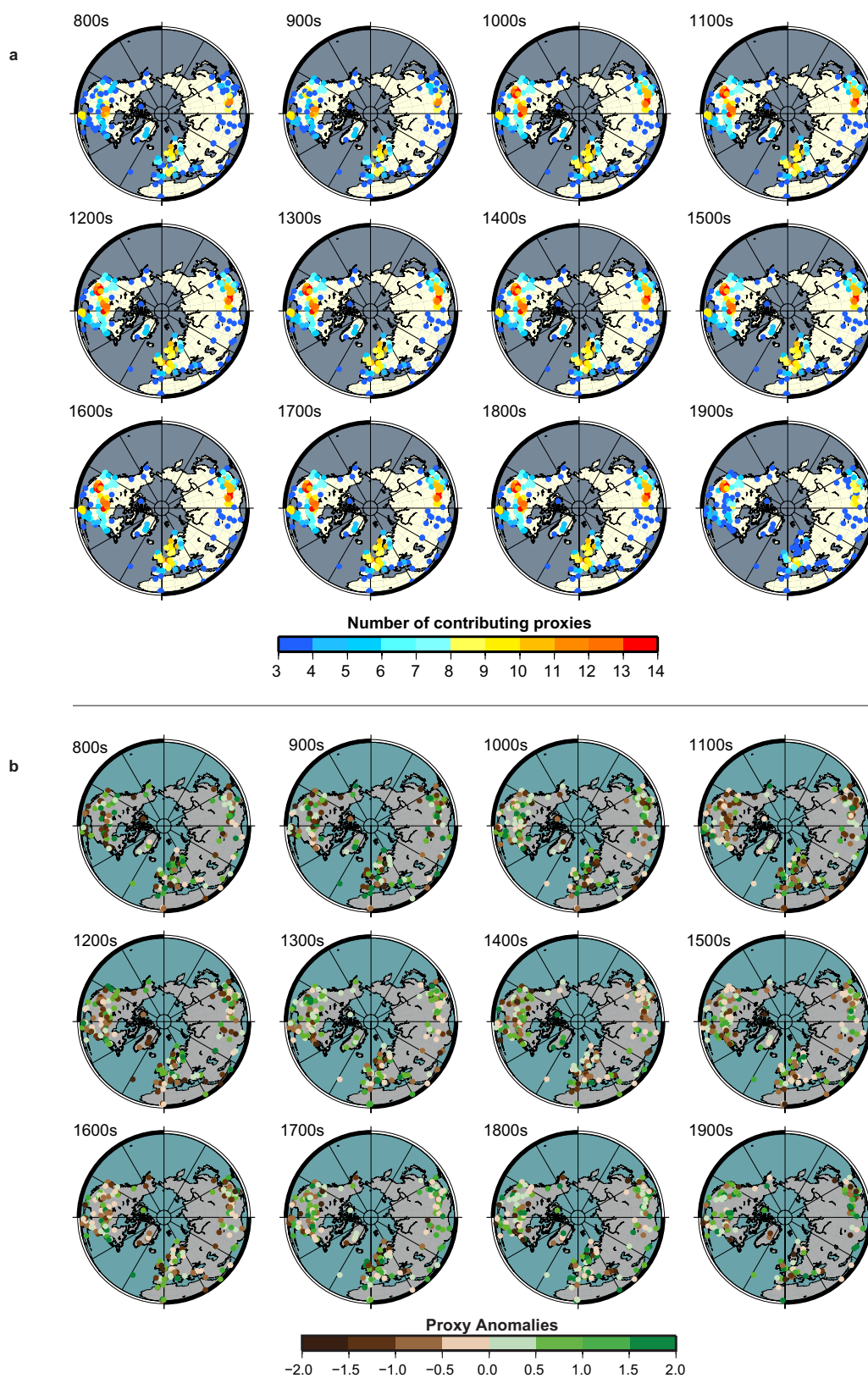
Extended Data Figure 6 | Centennial temperature proxy anomalies updated from ref. 15. a, Gridded, weighted, centennial proxy anomalies values derived from the data listed in Supplementary Table 2 and shown in Fig. 1b. Anomalies are shown relative to the centennial mean and standard deviation over the eleventh–nineteenth centuries. The colour scale is truncated at -2 and 2 and areas with insufficient proxy coverage

to compute a gridded weighted mean value are left white. **b,** Gridded, weighted, centennial anomalies for simulated median values of annual mean temperature from six atmosphere–ocean coupled general circulation models. Only values from the same grid cells that are covered by proxy records are extracted (Methods).



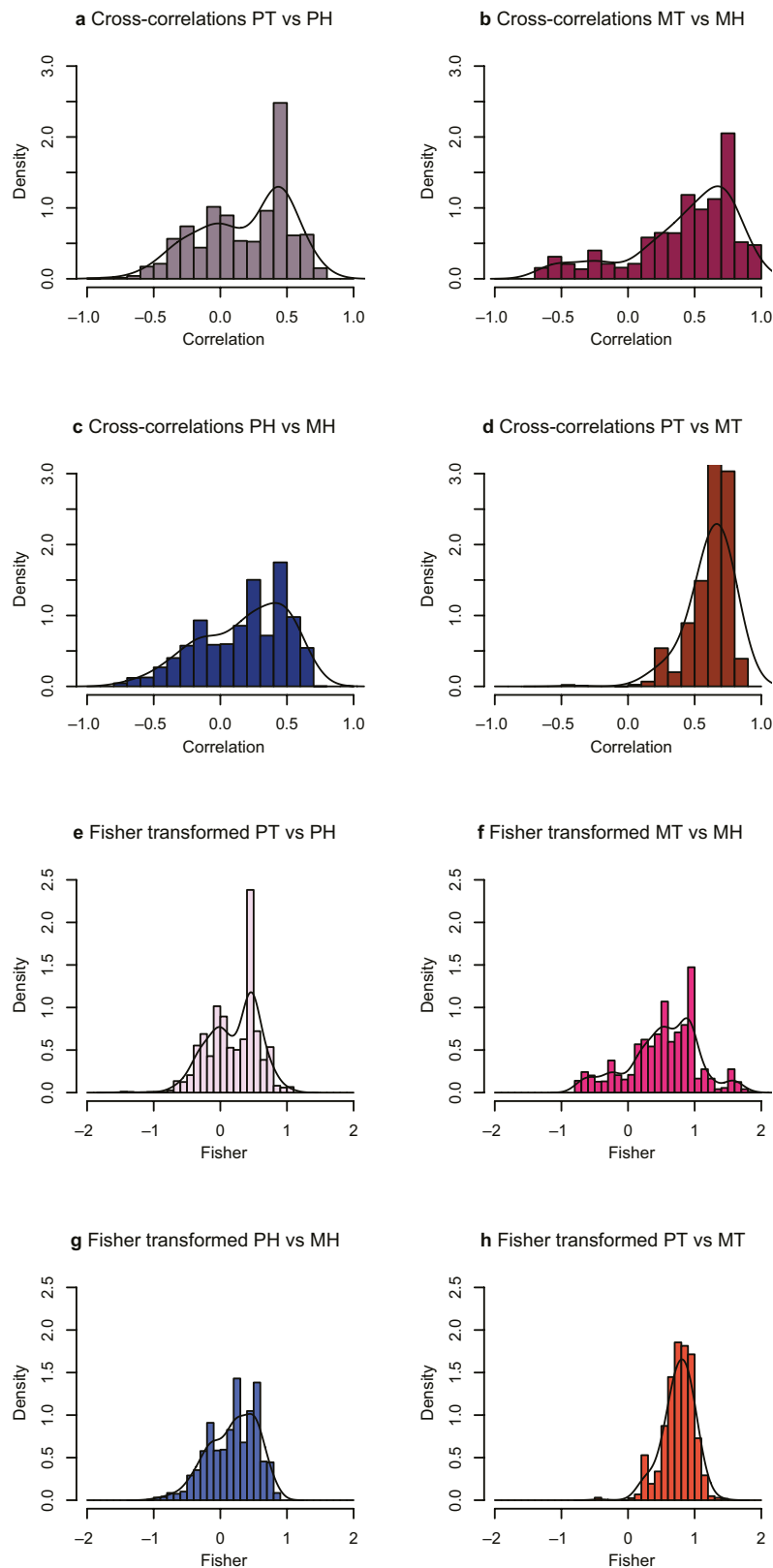
Extended Data Figure 7 | Gradients of proxy-reconstructed and simulated Northern Hemisphere centennial hydroclimate anomalies along three meridional transects for the tenth, twentieth and seventeenth centuries. The tenth and twentieth centuries were the warmest centuries of the past twelve and the seventeenth century was the coldest. **a**, Smoothed, surfaced, and contoured weighted average centennial proxy anomalies for the tenth century (top right). The trend of the smoothed surfaced anomaly values, with their regression line,

is shown along the meridional transects, passing through the densest data clusters (red line, North America; blue line, Europe and Africa; and black line, Asia). **b**, **c**, The same as **a** but depicting centennial proxy anomalies of hydroclimate for the seventeenth and the twentieth centuries, respectively. **d–f**, The equivalent analysis for the same centennial periods using median-model simulated values, extracted from the same proxy locations, of centennial precipitation anomalies (see Methods).



Extended Data Figure 8 | Distribution and density of hydroclimate proxy records. **a**, Number of contributing hydrological proxy records included in each proxy-centred anisotropic weighted mean calculation where there are three or more neighbouring proxies found in the search radius. **b**, Raw, centennial, hydroclimate proxy anomaly values derived

from the data listed in Supplementary Table 1 and shown in Fig. 1a. Anomalies are shown relative to the centennial mean and standard deviation over the eleventh–nineteenth centuries. The colour scale is truncated at -2 and 2 and areas with insufficient proxy coverage to compute a gridded weighted mean value are left white.



Extended Data Figure 9 | Histograms of cross-correlations with kernel density estimate added. **a**, Cross-correlations between gridded hydroclimate proxy data and temperature proxy (PT) data. **b**, Cross-correlations between gridded hydroclimate model data and temperature model data. **c**, Cross-correlations between gridded hydroclimate proxy (PH) data and hydroclimate model (MH) data. **d**, Cross-correlations

between gridded temperature proxy data and temperature model (MT) data. **e**, The same as **a** but Fisher-transformed. **f**, The same as **b** but Fisher-transformed. **g**, The same as **c** but Fisher-transformed. **h**, The same as **d** but Fisher-transformed. The thin curves in each histogram represent the kernel density estimates.

Extended Data Table 1 | Results from multiple sensitivity tests of our gridded weighted hydroclimate reconstruction

a	Median Pearson's correlation	Median <i>p</i>-value
Reconstruction from only records with at least decadal resolution and centennial or better dating control	0.97	< 0.01
Excluding ice cores	0.86	< 0.01
Excluding peat cores	0.90	< 0.01
Excluding tree rings	0.77	< 0.01
Excluding speleothems	0.84	< 0.01
Excluding documentary data	0.90	< 0.01
Excluding lake sediments	0.80	< 0.01
Excluding marine sediments	0.88	< 0.01
Retaining only calibrated precipitation records	0.89	< 0.01
Excluding calibrated precipitation records	0.92	< 0.01
Excluding records with annual signal	0.98	< 0.01
Excluding records with summer signal	0.77	< 0.01
Excluding records with winter signal	0.78	< 0.01
b		
North American Drought Atlas ²⁰	0.34	0.02
Monsoon Asia Drought Atlas ²¹	0.51	< 0.01
c		
GHCN5 instrumental annual precipitation data	0.66	0.05

a, Statistics from correlating the full gridded weighted reconstruction at the grid-cell level with reconstructions derived from using different subsets of proxy data including tests removing one proxy type at the time. **b**, Statistics from correlating the full gridded weighted reconstruction at the grid-cell level with PDSI drought atlas reconstructions. **c**, Results from correlating the decadal-resolved reconstruction from the proxy subset including only records with at least decadal resolution and centennial or better dating control with the decadal mean gridded 5° × 5° GHCN5 instrumental annual precipitation data over the twentieth century.

Late Quaternary climate change shapes island biodiversity

Patrick Weigelt^{1,2}, Manuel Jonas Steinbauer^{3,4}, Juliano Sarmiento Cabral^{1,5} & Holger Kreft¹

Island biogeographical models consider islands either as geologically static with biodiversity resulting from ecologically neutral immigration–extinction dynamics¹, or as geologically dynamic with biodiversity resulting from immigration–speciation–extinction dynamics influenced by changes in island characteristics over millions of years². Present climate and spatial arrangement of islands, however, are rather exceptional compared to most of the Late Quaternary, which is characterized by recurrent cooler and drier glacial periods. These climatic oscillations over short geological timescales strongly affected sea levels^{3,4} and caused massive changes in island area, isolation and connectivity⁵, orders of magnitude faster than the geological processes of island formation, subsidence and erosion considered in island theory^{2,6}. Consequences of these oscillations for present biodiversity remain unassessed^{5,7}. Here we analyse the effects of present and Last Glacial Maximum (LGM) island area, isolation, elevation and climate on key components of angiosperm diversity on islands worldwide. We find that post-LGM changes in island characteristics, especially in area, have left a strong imprint on present diversity of endemic species. Specifically, the number and proportion of endemic species today is significantly higher on islands that were larger during the LGM. Native species richness, in turn, is mostly determined by present island characteristics. We conclude that an appreciation of Late Quaternary environmental change is essential to understand patterns of island endemism and its underlying evolutionary dynamics.

The Quaternary is characterized by marked climatic oscillations between glacials and interglacials that had worldwide consequences for species distributions and diversity^{8,9}. Warm interglacials, like the Holocene (past 11,500 years), can be regarded as short-term anomalies embedded in much longer, cooler and drier glacial periods³. For

most of the Late Quaternary, massive amounts of water accumulated in continental ice sheets. This lowered sea levels considerably⁴ and influenced size, spatial arrangement and isolation of landmasses. In this context, the estimated 122 m sea level rise from the LGM (~21,000 years before present) to the early Holocene was particularly drastic and rapid³ (Fig. 1a).

Late Quaternary sea level lowstands had far-reaching consequences for island geography (Fig. 2). Many islands were much larger and less isolated^{5,10}. Neighbouring islands today separated by shallow waters formed single landmasses, for example, Fuerteventura and Lanzarote made up the larger Mahan¹¹ (Fig. 2d), and currently submerged seamounts emerged and created stepping-stones for species dispersal^{10–12}. Moreover, many islands were repeatedly connected and disconnected to the mainland^{13,14}. Such profound changes in island geography should have strongly affected immigration, speciation and extinction, the core processes determining the assembly and diversity of island biota^{5,7}. Present island diversity thus probably carries an imprint of past physical and bioclimatic conditions.

While the effects of Quaternary climate change on biodiversity are increasingly well-recognized for mainlands^{8,9,15}, they are under-represented in island biogeography theory^{5,7}. The equilibrium theory of island biogeography¹, perhaps the most influential framework in island biology, assumes species richness to result from a dynamic equilibrium between immigration (plus speciation) and extinction. Immigration and speciation rates are higher and extinction rates lower on larger islands, while immigration rates are lower and extinction and speciation rates higher on more isolated islands^{1,16–18}. The equilibrium theory assumes island geography not to change, and limitations of this static view have widely been recognized^{13,19}. More recent models emphasize the dynamic nature of islands and relate biogeographical rates and biodiversity to island ontogeny-related changes in area and topographical

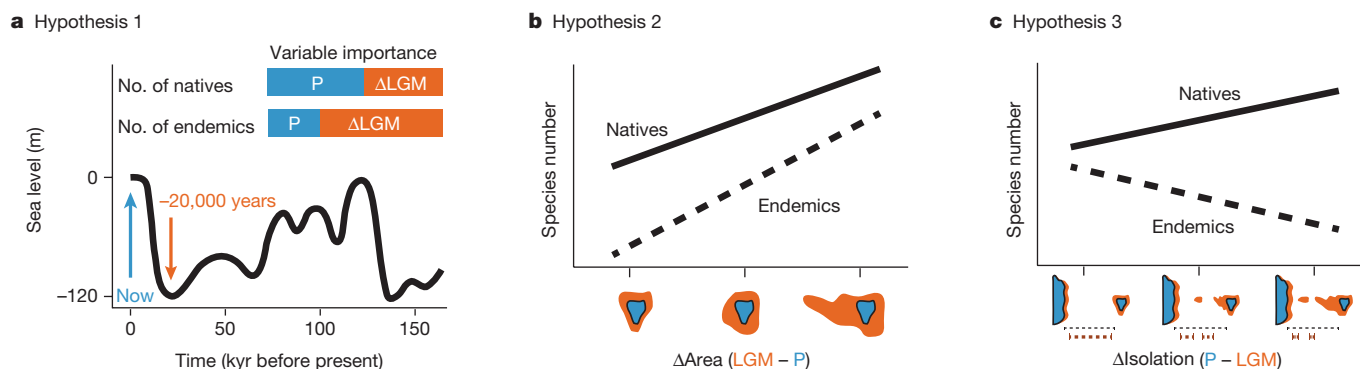


Figure 1 | Hypothesized effects of post-LGM sea-level changes on island biodiversity. **a**, We hypothesize that present (P) plant diversity has been affected by LGM island characteristics (sea-level curve in **a** after ref. 3). Changes in island characteristics since the LGM (ΔLGM) should therefore be important predictors of present diversity patterns. These effects should

be stronger for endemic than for native non-endemic species richness. kyr, thousand years. **b**, **c**, Hypothesized effects of changes in island area (ΔArea) (**b**) and stepping-stone isolation (Δisolation; considering also currently submerged seamounts and changes in mainland coastlines) (**c**) from LGM to present (P) on island plant diversity.

¹Biodiversity, Macroecology & Conservation Biogeography Group, University of Göttingen, 37077 Göttingen, Germany. ²Systemic Conservation Biology, University of Göttingen, 37073 Göttingen, Germany. ³Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, 8000 Aarhus, Denmark. ⁴Department of Biogeography, BayCEER, University of Bayreuth, 95440 Bayreuth, Germany. ⁵Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany.

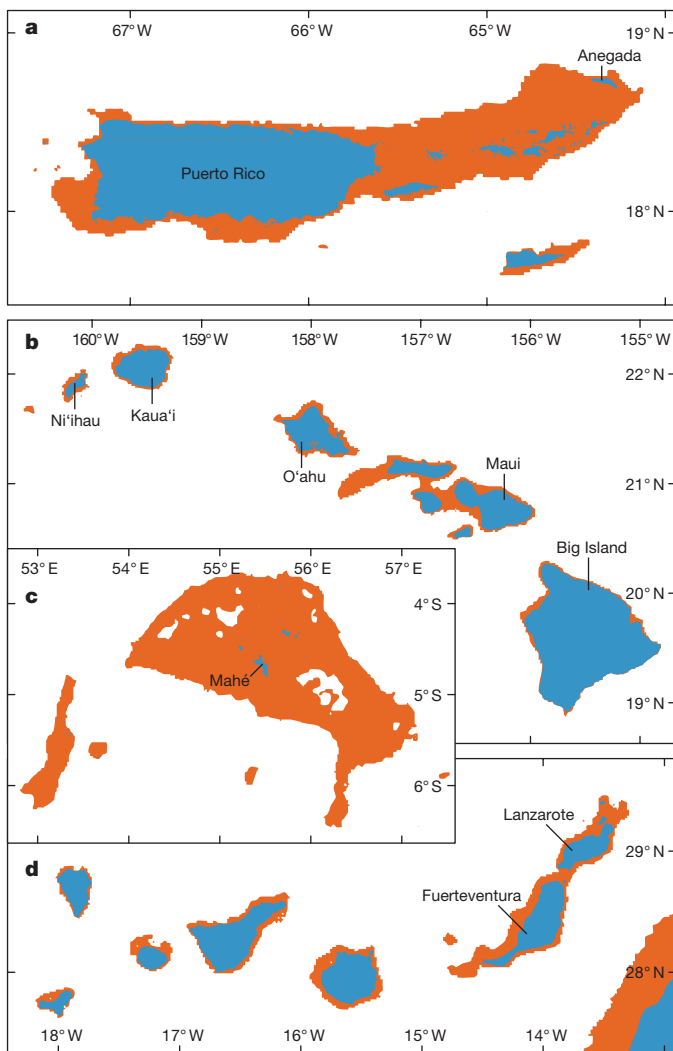


Figure 2 | Examples of spatial arrangements of present landmasses (blue) and modelled landmasses during the LGM (orange). LGM coastlines were estimated based on global bathymetry data³¹ assuming a sea-level decrease of 122 m at 20,000 years before present⁴ (see Methods). Present coastlines were taken from the GADM database of Global Administrative Areas (<http://www.gadm.org/version1>). **a**, Puerto Rico and British Virgin Islands. **b**, Hawaiian Islands. **c**, Seychelles. **d**, Canary Islands. Note the different changes in area, isolation and connectivity of islands. Exemplary islands mentioned in the text are named.

heterogeneity over long geological time scales^{2,6}. Also shallow-time changes (that is, changes over short geological time scales) in island area, isolation, elevation and climate caused by Late Quaternary oscillations, have recently been proposed to have influenced immigration, speciation and extinction⁵, but whether this has left an imprint in global island biodiversity remains untested.

Here, we evaluate present and shallow-time past drivers of species richness and endemism of angiosperms (flowering plants) on 184 islands worldwide. We hypothesize post-LGM changes in physical (area, isolation, elevation) and bioclimatic (temperature, precipitation) island characteristics to be strong predictors of present angiosperm diversity (hypothesis 1; Fig. 1a). Especially endemic species richness, which relates to evolutionary dynamics, should reflect past physical and bioclimatic conditions⁹. Considering that islands were larger during the LGM, islands with greater area change (Δ area) should harbour both more native and endemic species than predicted by their present size (hypothesis 2; Fig. 1b). Islands with greater changes in isolation (Δ isolation) should harbour more native, but fewer endemic species (hypothesis 3; Fig. 1c), owing to higher LGM connectivity of

landmasses increasing inter-island dispersal as well as colonization from and to the mainland.

We estimated LGM coastlines of islands and continents assuming a eustatic sea level drop of 122 m (ref. 4). The 184 studied islands experienced strong post-LGM changes in most physical and bioclimatic conditions owing to warming climate and resulting sea level rise (Fig. 2 and Extended Data Fig. 1). Island area decreased by 5,035 km² and 71.1% on average (ranging from −4 to −61,377 km²; −6.3% to −99.997%). Islands that are currently small, but merged with other islands during the LGM, experienced particularly strong proportional reductions in area (for example, −99.8% for Anegada, British Virgin Islands; −99.6% for Mahé, Seychelles; Fig. 2a, c). These post-LGM changes have been much faster than average rates of geologic processes of island formation, subsidence, and erosion (Extended Data Fig. 2). Island isolation measured as distance to the nearest landmass >100,000 km² accounting for present stepping-stone islands¹² and currently submerged seamounts increased on average by 107 km (+3 to +513 km; Extended Data Fig. 1). Annual temperatures increased on average by 2 °C and annual precipitation by 163 mm changing in both directions (−2,836 to +2,684 mm). Changes in elevation were high if a focal island was connected to a high-elevation island (+327 m on average).

Regression analyses showed that present island biodiversity is strongly related to post-LGM changes in island characteristics (Fig. 3), supporting hypothesis 1. Endemism measured as the number and proportion of endemic species, in particular, was strongly affected by LGM-related variables (sum of scaled Akaike's information criterion-weighted standardized regression coefficients >0.5; Fig. 3a) and could be well explained by a combination of post-LGM changes and present island characteristics (adjusted deviance explained (D^2_{adj}) all >0.8; Extended Data Table 1). Effects of area change via sea-level change on endemism were strong, whereas direct effects of changes in temperature and precipitation were weak. Species richness, in contrast, was mostly related to present environmental characteristics, and post-LGM changes showed only minor effects (Fig. 3a). The strong effects of present environments on species richness might be explained by diversity dynamics closely following climate-induced changes²⁰, suggesting that overall species richness on islands is close to equilibrium with present conditions.

Island area was the variable that changed most since the LGM, and the correlation between LGM and present area was weak ($r=0.32$, $P<0.001$; Extended Data Fig. 3). Area is a key variable in island biogeography: first, larger islands sustain larger populations resulting in lower extinction rates¹; second, larger islands receive more colonizers via target-area effects¹⁷; third, large islands have higher within-island speciation rates^{16,21,22} owing to more heterogeneous environments (correlation between area and elevation: $r=0.74$, $P<0.001$), which is supported here by strong effects of present elevation on number and proportion of endemic species²³ (Fig. 3a and Extended Data Table 1). Moreover, large, high-elevation islands may be refugia for once widespread lowland species and montane species that became extinct on lower islands^{5,9}. Accordingly, species richness and particularly richness and proportion of endemics were higher on islands that were larger during the LGM (Fig. 3b and Extended Data Fig. 4), supporting hypothesis 2. However, the positive effect of Δ area on species richness was small (Fig. 3b), indicating that the number of species on a shrunken island is only slightly larger than expected from its present area. Considering this and that island floras might have been unsaturated during the LGM²⁴, a potential post-LGM extinction debt²⁵ seems to have a minor role.

The effect of Δ area on native species richness emerged mainly from a strong effect of Δ area on endemic richness (single-island endemics (SIE) in Fig. 3 and Extended Data Table 1). With endemic species excluded, in fact, the slope of the Δ area–richness relationship was even shallower (standardized regression coefficient and variable importance of Δ area: endemics > natives > native non-endemics; Fig. 3a, b). Island endemics usually result from *in situ* differentiation rather than

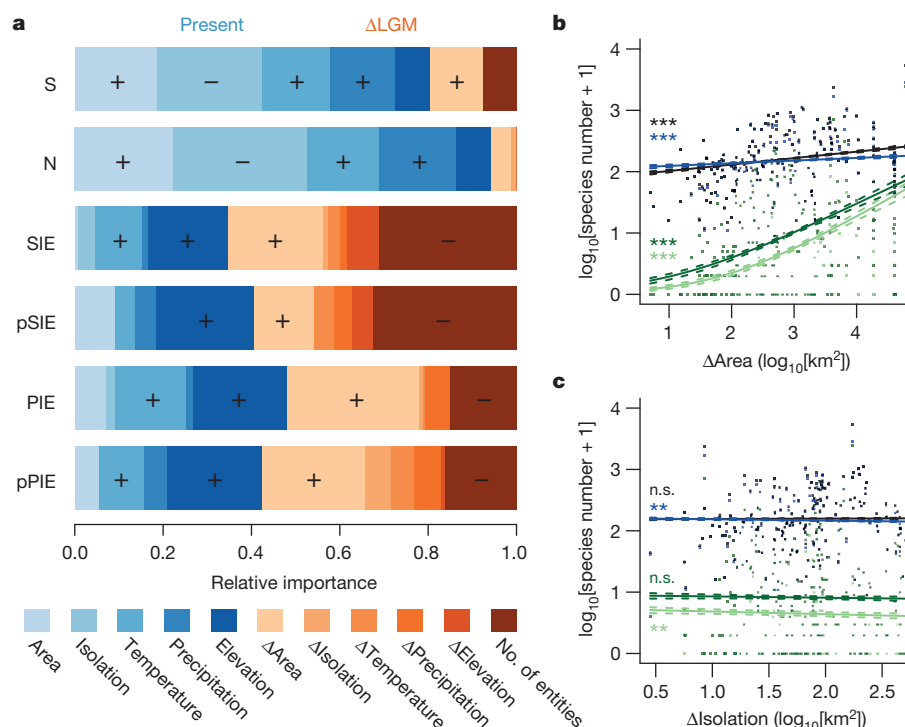


Figure 3 | Predictors of angiosperm diversity on 184 islands worldwide. **a**, Relative importance of present island characteristics (blue) and post-LGM changes (Δ LGM; orange) as predictors of angiosperm diversity, that is, number of native species (S), native non-endemics (N), single-island endemics (SIE), species endemic to past island units (PIE), and proportions of single-island endemics (pSIE) and species endemic to past islands (pPIE). Δ LGM variables quantify the difference in island characteristics between LGM and today. No. of entities, number of present

island entities that were part of single LGM islands; +/–, direction of important effects. **b**, **c**, Effects of post-LGM changes in island area (Δ Area) (**b**) and stepping-stone isolation (Δ Isolation) (**c**) controlling for all covariables in generalized linear models. Relationships and 95% confidence intervals (dashed) are shown for S (black), N (blue), SIE (light green) and PIE (green). ** $P < 0.01$, *** $P < 0.001$, NS, not significant ($P \geq 0.05$).

from extinction elsewhere¹³. Islands repeatedly larger in the past may hence have been arenas of within-island speciation causing a strong imprint in present endemism patterns. Consequently, endemic species may inhabit islands presently too small for frequent within-island isolation and associated *in situ* differentiation¹⁶. A further explanation for the higher number and proportion of endemics on formerly larger islands would be a lower extinction risk of endemics relative to native non-endemic species. Species evolved on islands might be pre-adapted to smaller population sizes and dynamic island environments, and might be less affected by decreasing area, which is in contrast to previous interpretations²⁶.

Contrary to hypothesis 3, Δ isolation had no effect on native and endemic species richness and the proportion of endemics (Fig. 3a, c and Extended Data Fig. 4). The difference between present and LGM island isolation was small and both were strongly related ($r = 0.997$, $P < 0.001$; Extended Data Fig. 3). Furthermore, most islands in our study are quite isolated from the mainland (mean = 1,427 km) and consequently immigration events are generally rare. For Hawaiian plants, for instance, the net colonization rate (disregarding extinction) is estimated at one colonizer every 98,000 years²⁷. Sea-level changes may therefore have only weakly affected rates of long-distance dispersal from major landmasses, even though isolation undoubtedly affects immigration¹ and speciation².

Probably more importantly, inter-island dispersal was affected by post-LGM changes in within-archipelago connectivity^{11,28}. Accordingly, we found a strong negative effect of the number of present islands that merged during the LGM on the number of single-island endemics (SIE in Fig. 3a and Extended Data Fig. 5). This means that islands that only lost area since the LGM have on average more single-island endemics than similar sized-islands that were connected to other islands, indicating a strong role of past connectivity for intra-archipelagic dispersal.

Inter-island dispersal of single-island endemics immediately changes their endemism status leading to fewer single-island endemic species on formerly connected islands. Accordingly, when considering species endemic to the past island entities, the effect of the number of present islands decreased (compare past-island endemics (PIE) in Fig. 3a and Extended Data Fig. 5). However, the still significant negative effect contradicts the notion of repeated fusion and fission of islands acting as speciation pump²², which might at the given spatial and temporal scale only be important for certain taxonomic groups of low dispersal ability (compare ref. 10). Dispersal and gene flow between formerly connected islands appears to impede rather than promote speciation in plants. The strong positive effect of Δ area on endemic richness hence suggests that speciation is promoted in cases of high environmental opportunity and intra-archipelagic connectivity without actual land-bridge connections. Considering these area changes and fusion–fission dynamics may also help to understand additional biogeographical patterns and processes, for example, why certain lineages or functional groups diversify more frequently on islands than others²⁹.

In conclusion, sea-level changes during the Late Quaternary have left a strong imprint on the present distribution of island endemics. Especially, the strong effects of changes in area and intra-archipelagic connectivity on endemism suggest a heretofore underestimated legacy of LGM conditions in immigration, speciation and extinction dynamics on islands. Island biogeography theory hence needs to incorporate the effects of shallow-time geographic dynamics on inter-island dispersal and diversification^{2,5}. Future island research should also focus on other, poorly understood LGM-associated shifts in, for example, ocean currents, wind systems, or source areas. Considering the disproportional contribution of island endemics to global plant diversity³⁰, these dynamics have profound implications for global biodiversity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 December 2015; accepted 23 February 2016.

Published online 30 March; corrected online 6 April 2016

(see full-text HTML version for details).

1. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton University Press, 1967).
2. Whittaker, R. J., Triantis, K. A. & Ladle, R. J. A general dynamic theory of oceanic island biogeography. *J. Biogeogr.* **35**, 977–994 (2008).
3. Bintanja, R., van de Wal, R. S. & Oerlemans, J. Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* **437**, 125–128 (2005).
4. Miller, K. G. *et al.* The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298 (2005).
5. Fernández-Palacios, J. M. *et al.* Towards a glacial-sensitive model of island biogeography. *Glob. Ecol. Biogeogr.* (2015).
6. Borregaard, M. K. *et al.* Oceanic island biogeography through the lens of the general dynamic model: assessment and prospect. *Biol. Rev. Camb. Philos. Soc.* (2016).
7. Warren, B. H. *et al.* Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
8. Dynesius, M. & Jansson, R. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proc. Natl Acad. Sci. USA* **97**, 9115–9120 (2000).
9. Sandel, B. *et al.* The influence of Late Quaternary climate-change velocity on species endemism. *Science* **334**, 660–664 (2011).
10. Ali, J. R. & Aitchison, J. C. Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *J. Biogeogr.* **41**, 1227–1241 (2014).
11. Rijdsdijk, K. F. *et al.* Quantifying surface-area changes of volcanic islands driven by Pleistocene sea-level cycles: biogeographical implications for the Macaronesian archipelagos. *J. Biogeogr.* **41**, 1242–1254 (2014).
12. Weigelt, P. & Kreft, H. Quantifying island isolation – insights from global patterns of insular plant species richness. *Ecography* **36**, 417–429 (2013).
13. Heaney, L. R. Dynamic disequilibrium: a long-term, large-scale perspective on the equilibrium model of island biogeography. *Glob. Ecol. Biogeogr.* **9**, 59–74 (2000).
14. Weigelt, P., Jetz, W. & Kreft, H. Bioclimatic and physical characterization of the world's islands. *Proc. Natl Acad. Sci. USA* **110**, 15307–15312 (2013).
15. Svenning, J.-C. & Skov, F. Ice age legacies in the geographical distribution of tree species richness in Europe. *Glob. Ecol. Biogeogr.* **16**, 234–245 (2007).
16. Kisel, Y. & Barraclough, T. G. Speciation has a spatial scale that depends on levels of gene flow. *Am. Nat.* **175**, 316–334 (2010).
17. Whitehead, D. R. & Jones, C. E. Small islands and the equilibrium theory of insular biogeography. *Evolution* **23**, 171–179 (1969).
18. Brown, J. H. & Kodric-Brown, A. Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology* **58**, 445–449 (1977).
19. Ricklefs, R. E. & Bermingham, E. Nonequilibrium diversity dynamics of the Lesser Antillean avifauna. *Science* **294**, 1522–1524 (2001).
20. Nogué, S. *et al.* The ancient forests of La Gomera, Canary Islands, and their sensitivity to environmental change. *J. Ecol.* **101**, 368–377 (2013).
21. Losos, J. B. & Schluter, D. Analysis of an evolutionary species-area relationship. *Nature* **408**, 847–850 (2000).
22. Gillespie, R. G. & Roderick, G. K. Evolution: geology and climate drive diversification. *Nature* **509**, 297–298 (2014).
23. Steinbauer, M. J., Irl, S. D. H. & Beierkuhnlein, C. Elevation-driven ecological isolation promotes diversification on Mediterranean islands. *Acta Oecol.* **47**, 52–56 (2013).
24. Stohlgren, T. J., Barnett, D. T., Jarnevich, C. S., Flather, C. & Kartesz, J. The myth of plant species saturation. *Ecol. Lett.* **11**, 313–322 (2008).
25. Jackson, S. T. & Sax, D. F. Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover. *Trends Ecol. Evol.* **25**, 153–160 (2010).
26. Triantis, K., Mylonas, M. & Whittaker, R. Evolutionary species–area curves as revealed by single-island endemics: insights for the inter-provincial species–area relationship. *Ecography* **31**, 401–407 (2008).
27. Price, J. P. & Clague, D. A. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proc. R. Soc. Lond. B* **269**, 2429 (2002).
28. Cabral, J. S., Weigelt, P., Kissling, W. D. & Kreft, H. Biogeographic, climatic and spatial drivers differentially affect α -, β - and γ -diversities on oceanic archipelagos. *Proc. R. Soc. Lond. B* **281**, 20133246 (2014).
29. Price, J. P. & Wagner, W. L. A phylogenetic basis for species–area relationships among three Pacific Island floras. *Am. J. Bot.* **98**, 449–459 (2011).
30. Kier, G. *et al.* A global assessment of endemism and species richness across island and mainland regions. *Proc. Natl Acad. Sci. USA* **106**, 9322–9327 (2009).
31. Amante, C. & Eakins, B. W. ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. *NOAA Technical Memorandum NESDIS NGDC-24* (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements P.W., J.S.C. and H.K. acknowledge funding by the German Research Council (DFG) Free Floater Program in the Excellence Initiative at the University of Göttingen. P.W. and H.K. additionally acknowledge funding by the BEFmate project from the Ministry of Science and Culture of Lower Saxony. M.J.S. was supported by the Danish Carlsbergfondet (CF14-0148). We are grateful to S. L. Chown, J. Gerlach, Y. Kisel, J. P. Price and J. D. Shaw for providing species lists. We thank R. J. Whittaker for helpful discussions and J.-C. Svenning for comments on a previous version of the manuscript.

Author Contributions All authors designed the study. P.W. and H.K. collected the data. P.W. led the analyses with contributions of M.J.S. and J.S.C. All authors jointly wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.W. (pweigel@uni-goettingen.de) or H.K. (hkreft@uni-goettingen.de).

METHODS

Island data. We estimated last glacial maximum (LGM) coastlines of islands and mainlands based on ETOPO1 global bathymetry data³¹ at 1-arcminute resolution assuming a eustatic sea level decrease of 122 m at 20,000 years before present⁴ (Fig. 2). High-resolution present coastlines were taken from the GADM database of Global Administrative Areas, version 1 (<http://www.gadm.org/version1>). Past and present island area and isolation were derived from the respective coastline data (Fig. 2). To account for present stepping-stone islands as well as seamounts that emerged as potential stepping-stones during the LGM, we used the shortest distance over water from an island to the nearest landmass larger than 100,000 km² as isolation metric (metric stD₉ in ref. 12). Calculating isolation as the proportion of water in the surrounding of each island as suggested by ref. 12 led to lower values of present isolation compared to the LGM when large neighbouring islands were merged in the past (Extended Data Fig. 3). We therefore opted for using stepping-stone isolation as isolation metric. In addition, the number of present island entities covered by species checklists that formed part of LGM island units entered the analyses as a measure of past inter-island connectivity (no. of entities). Island elevation of both present and past islands was derived from a global digital elevation model³². For past islands, elevation was the maximum elevation of all present islands included in the past island unit plus 122 m.

During the past one million years, sea levels repeatedly fell below –100 m and today's sea level represents one of few highstands⁴. The island characteristics calculated for the LGM are hence proxies for recurrent conditions that might have shaped today's island biotas over at least one million years. In addition to global eustatic sea-level changes, oceanic island subsidence and loading of the water column may affect regional sea levels and therefore also the area, isolation and elevation of oceanic islands at the relevant time scale of millennia^{10,11}. At the global scale, however, these effects are currently difficult to address, and, compared to the magnitude of eustatic sea-level changes, their effects are small^{10,11}. Eustatic fluctuations are hence the most relevant driver of recurrent sea-level changes at the global extent and at the relevant time scale (compare Extended Data Fig. 2).

Maximum values of present annual mean temperature and annual precipitation per island were derived from WorldClim^{14,32}. For past climatic variables, hindcasts for 21,000 years before present were derived from the CCSM3 and MIROC3.2 models³³, downscaled to 30 arcseconds. We averaged the two model predictions as LGM mean annual temperature and precipitation according to refs 9, 14. Maximum values of past annual mean temperature and annual precipitation were extracted based on all present islands belonging to a past island unit. Note that climate hindcasts for islands are considered less reliable compared to mainland hindcasts³⁴. We therefore focus our interpretations on the physical island characteristics area, isolation and number of island entities merged in the past and only consider climatic changes as covariables. In addition to natural changes in island characteristics, human-induced changes since the LGM influenced island floras³⁵. However, here we only considered native species and whenever possible included species that recently went extinct owing to human activity to focus on natural drivers of island plant diversity.

For 184 oceanic islands worldwide that were not connected to the mainland during the LGM¹⁴, we derived angiosperm species richness of native species (S), native non-endemic species (N), species endemic to single present island entities (SIE), and species endemic to all islands that were comprised by single island entities during the LGM (PIE), based on a global database of island floras^{36,37} (Extended Data Fig. 6). No statistical methods were used to predetermine sample size. The subset of the database used here comprised species lists from published floras, checklists and online databases^{38–91} as well as data from original research papers provided by the authors^{16,29,92,93}. Infra-specific endemics were not considered. As native non-endemic species, we considered species neither endemic to present nor to past island entities (for correlations among diversity metrics see Extended Data Fig. 7). Islands were included only if all islands of the past island unit were covered by species lists and islands were excluded if their checklists referred to more than one past island unit (for example, Puerto Rico). In total, the data set included 17,626 species including 6,257 island endemics. All species names were subjected to a comprehensive taxonomic standardization based on The Plant List (version 1.1, <http://www.theplantlist.org>) and iPlant's taxonomic name resolution service (version 3.2, <http://tnrs.iplantcollaborative.org>), following ref. 36.

Statistical analysis. We used present island characteristics and the differences between past and present conditions (LGM values – present values, depicted by Δ) as predictor variables for plant diversity and endemism. Because isolation and temperature increased since the LGM, we multiplied Δ isolation and Δ temperature by –1 to be able to relate increases in environmental change to the diversity metrics. Area, precipitation, elevation, Δ area, Δ isolation and Δ elevation were log₁₀-transformed after adding 1 in the case of elevation and Δ area to normalize their distributions and to moderate differences between very large values compared to differences between smaller values. Present island characteristics were only weakly correlated

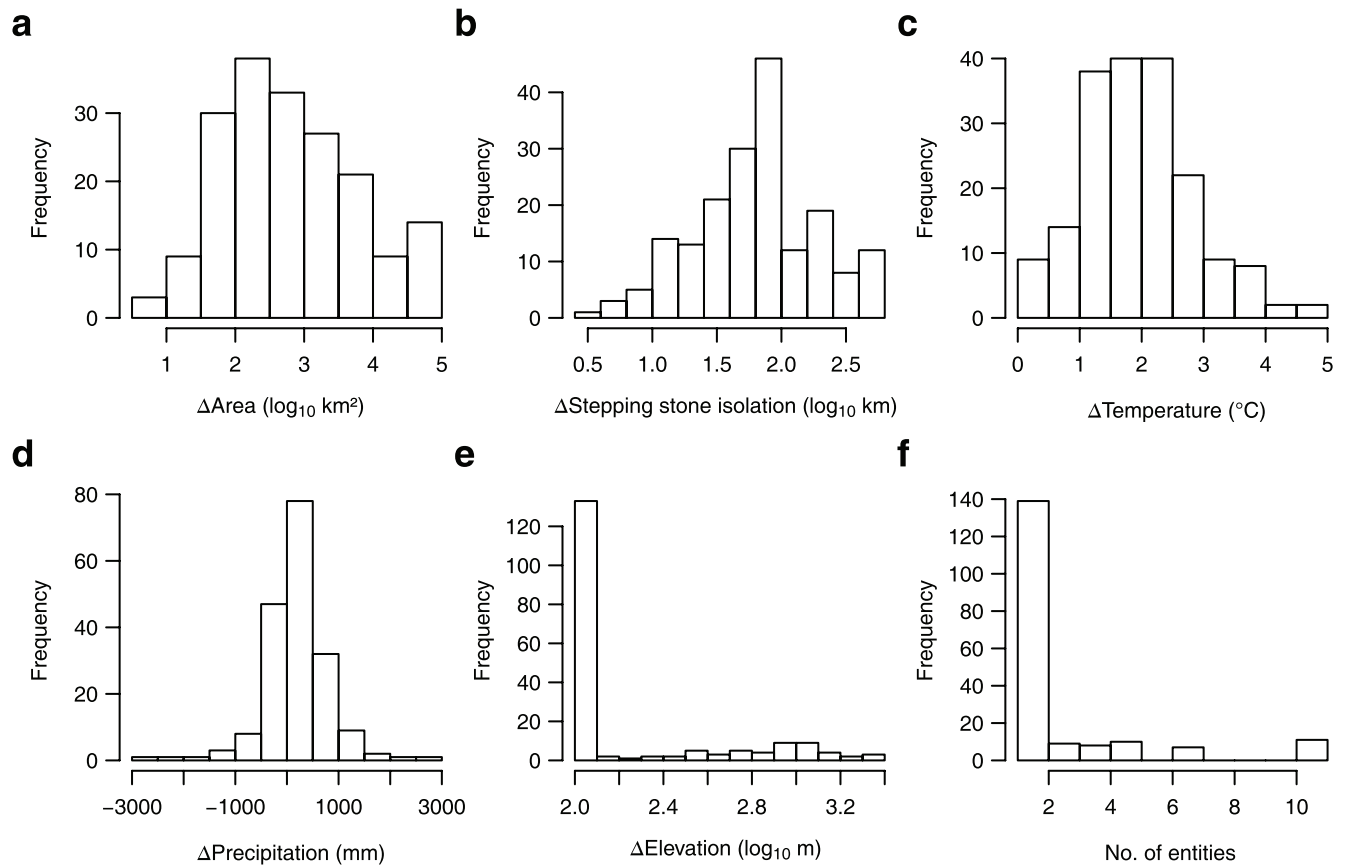
to the respective differences between past and present (all $|r| < 0.44$; Extended Data Table 2).

We modelled species numbers (S, N, SIE, PIE) using generalized linear models with a Poisson distribution. For proportions of endemic species (pSIE, pPIE), we used models with a binomial distribution where endemic species and native non-endemic species entered as composite response variables to account for differences in species richness. Since spatial autocorrelation in model residuals might violate the assumptions of generalized linear models⁹⁴, we included a spatial autocovariate built from residuals of the non-spatial model⁹⁵. The autocovariate was computed using the function `autocov_dist` from the R package `spdep`⁹⁶ using an optimized spatial neighbourhood structure with a neighbourhood distance of 1,000 km^{12,97}. We calculated models with all possible predictor variable combinations and selected the best models from all candidate models based on Akaike's information criterion (AIC) (Extended Data Table 1)⁹⁸. For graphical representation of the relationships between post-LGM changes and diversity metrics, we used full-model predictions including all predictor variables (Fig. 3b, c and Extended Data Figs 4 and 5). Regression lines were estimated with covariables held constant at the mean of their empirical values across the islands. As a measure of variable importance, we used AIC-weighted standardized regression coefficients. The standardized regression coefficients of all candidate models were weighted by the models' AIC weights and summed up for all possible models including a given variable^{28,98}. The variable importance values of all predictor variables were then scaled to sum up to one for comparison across models of richness of native and endemic species and proportions of endemic species (Fig. 3a).

Analyses were performed using ArcGIS 9.3 (ESRI, Redlands, USA) and the R statistical programming language (version 3.1.2, R Core Team). All diversity metrics, present bioclimatic and physical characteristics as well as post-LGM changes as used in the statistical analyses can be found in the Supplementary Data.

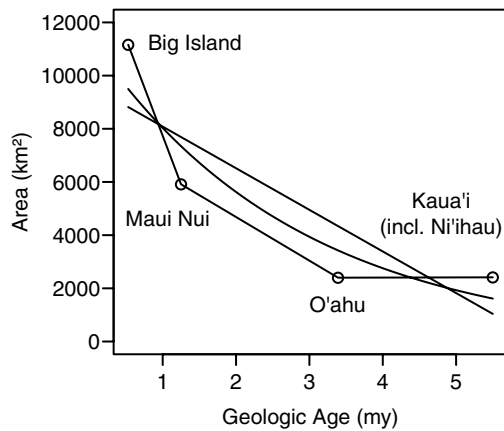
32. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
33. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the mid-Holocene and Last Glacial Maximum – part 1: experiments and large-scale features. *Clim. Past* **3**, 261–277 (2007).
34. Harter, D. E. V. *et al.* Impacts of global climate change on the floras of oceanic islands - Projections, implications and current knowledge. *Perspect. Plant Ecol. Evol. Syst.* **17**, 160–183 (2015).
35. Sax, D. F., Gaines, S. D. & Brown, J. H. Species invasions exceed extinctions on islands worldwide: a comparative study of plants and birds. *Am. Nat.* **160**, 766–783 (2002).
36. Weigelt, P. *et al.* Global patterns and drivers of phylogenetic structure in island floras. *Sci. Rep.* **5**, 12213 (2015).
37. Weigelt, P. The macroecology of island floras. *Frontiers of Biogeography* **7**, 119–125 (2015).
38. Abe, T. Threatened pollination systems in native flora of the Ogasawara (Bonin) Islands. *Ann. Bot. (Lond.)* **98**, 317 (2006).
39. Acevedo-Rodríguez, P. & Strong, M. T. *Catalogue of the seed plants of the West Indies* <http://botany.si.edu/antilles/WestIndies/index.htm>
40. Arechavaleta, M., Zurita, N., Marrero, M. C. & Martín, J. L. *Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres)* (Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, 2005).
41. Arechavaleta, M., Rodríguez, S., Zurita, N. & García, A. *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres* (Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, 2009).
42. Ashmole, P. & Ashmole, M. *St Helena and Ascension Island: a Natural History* (Anthony Nelson Ltd, 2000).
43. Athens, J. S., Blinn, D. W. & Ward, J. V. Vegetation history of Laysan Island, northwestern Hawaiian islands. *Pac. Sci.* **61**, 17–37 (2007).
44. Baker, M. L. & Duretto, M. F. *A Census of The Vascular Plants of Tasmania* (Tasmanian Herbarium, Tasmanian Museum and Art Gallery, 2011).
45. Borges, P. A. V. *et al.* *Listagem dos fungos, flora e fauna terrestres dos arquipélagos da Madeira e Selvagens* (Direcção Regional do Ambiente da Madeira and Universidade dos Açores, 2008).
46. Borges, P. A. V. *et al.* *A List of the Terrestrial and Marine Biota from the Azores* (Principia, Cascais, 2010).
47. Brofas, G., Karetos, G., Panitsa, M. & Theocharopoulos, M. The flora and vegetation of Giali Island, SE Aegean, Greece. *Willdenowia* **31**, 51–70 (2001).
48. Broughton, D. A. & McAdam, J. H. A checklist of the native vascular flora of the Falkland Islands (Islas Malvinas): new information on the species present, their ecology, status and distribution. *J. Torrey Bot. Soc.* **132**, 115–148 (2005).
49. Burton, F. J. *Red List Assessment of Cayman Islands' Native Flora for Legislation and Conservation Planning* (Royal Botanic Gardens Kew, 2007).
50. Caribbean Research and Management of Biodiversity Foundation. *Dutch Caribbean Biodiversity Explorer* <http://www.dcbiodata.net/explorer/home>
51. Case, T. J., Cody, M. L. & Ezcurra, E. *A new island biogeography of the Sea of Cortés* (Oxford University Press, 2002).
52. Charters, M. *Flora of Bermuda* <http://www.calflora.net/floraofbermuda/index.html>

53. Christmas Island National Park. *Third Christmas Island National Park Management Plan* (Parks Australia North, Christmas Island, Australia, 2002).
54. Cronk, Q. C. B. The past and present vegetation of St Helena. *J. Biogeogr.* **16**, 47–64 (1989).
55. D'Arcy, W. G. The island of Anegada and its flora. *Atoll Res. Bull.* **139**, 1–21 (1971).
56. de Lange, P. J., Heenan, P. B. & Rolfe, J. R. *Checklist of Vascular Plants Recorded from Chatham Islands* (Department of Conservation, Wellington Hawke's Bay Conservancy, 2011).
57. de Miranda Freitas, A. M. *A Flora Fanerogâmica Atual do Arquipélago de Fernando de Noronha - Brasil* (Universidade Estadual, 2007).
58. Du Puy, D. J. *Christmas Island: species lists* <http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/index.html>
59. Exell, A. W. *Catalogue of the Vascular Plants of S. Tome (with Principe and Annobon)* (Trustees of the British Museum, 1944).
60. Florence, J., Chevillotte, H., Ollier, C. & Meyer, J.-Y. *Base de données botaniques Nadeaud de l'Herbier de la Polynésie française (PAP)* <http://www.herbier-tahiti.pf>
61. Fosberg, F. R., Renvoize, S. A. & Townsend, C. C. *The flora of Aldabra and neighbouring islands* (HMSO, 1980).
62. Fosberg, F. R. & Sachet, M. H. Flora of Maupiti, Society Islands. *Atoll Res. Bull.* **294**, 1–70 (1987).
63. Green, P. S. *Lord Howe Island: species lists* <http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/49/index.html>
64. Greene, S. & Walton, D. An annotated check list of the sub-Antarctic and Antarctic vascular flora. *Polar Rec. (Gr. Brit.)* **17**, 473–484 (1975).
65. Hill, M. J. *Biodiversity Surveys and Conservation Potential of Inner Seychelles Islands* (Smithsonian Institution, 2002).
66. Hnatiuk, R. J. *Subantarctic Islands: species lists* <http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/index.html>
67. Imada, C. T. Hawaiian Native and Naturalized Vascular Plants Checklist. *Bishop Museum Technical Report* **60** (2012).
68. Jaramillo Diaz, P. & Guézou, A. *CDF checklist of Galapagos vascular plants* <http://www.darwinfoundation.org/datazone/checklists/vascular-plants/>
69. Johnson, P. N. & Campbell, D. J. Vascular plants of the Auckland Islands. *N.Z. J. Bot.* **13**, 665–720 (1975).
70. Johnston, I. M. The flora of the Revillagigedo Islands. *Proc. Calif. Acad. Sci.* **20**, 9–104 (1931).
71. Junak, S., Philbrick, R., Chaney, S. & Clark, R. *A Checklist of Vascular Plants of Channel Islands National Park*. 2nd edn (Southwest Parks and Monuments Association, 1997).
72. Kingston, N., Waldren, S. & Bradley, U. The phytogeographical affinities of the Pitcairn Islands – a model for south-eastern Polynesia? *J. Biogeogr.* **30**, 1311–1328 (2003).
73. Kirchner, F., Picot, F., Merceron, E. & Gigot, G. *Flore vasculaire de La Réunion* (Conservatoire Botanique National de Mascarin, 2010).
74. Levin, G. A. & Moran, R. The vascular flora of Socorro, Mexico. *Memoirs of the San Diego Society of Natural History* **16**, 1–71 (1989).
75. Marticorena, C., Stuessy, T. F. & Baeza, C. M. Catalogue of the vascular flora of the Robinson Crusoe or Juan Fernández islands, Chile. *Gayana Botanica* **55**, 187–211 (1998).
76. Miller, A. G. & Morris, M. *Ethnoflora of the Soqatra Archipelago* (Royal Botanic Garden, 2004).
77. Moran, R. *The Flora of Guadalupe Island, Mexico* (California Academy of Sciences, 1996).
78. Renvoize, S. A. A floristic analysis of the western Indian Ocean coral islands. *Kew Bull.* **30**, 133–152 (1975).
79. Sáez, L. & Rosselló, J. A. *Llibre vermell de la flora vascular de les Illes Balears* (Direcció General de Biodiversitat, Conselleria de Medi Ambient, Govern de les Illes Balears, 2001).
80. St John, H. *Census of the Flora of the Gambier Islands, Polynesia: Pacific Plant Studies* **43** (1988).
81. Strahm, W. A. *The Conservation and Restoration of the Flora of Mauritius and Rodrigues* (University of Reading, 1993).
82. Sykes, W. R. *Contributions to the Flora of Niue*. Department of Scientific and Industrial Research, New Zealand. Bulletin; 200 (1970).
83. Takahashi, H. et al. A preliminary checklist of the vascular plants of Chirinkotan, Kuril Islands. *Journal of Phytogeography and Taxonomy* **47**, 131–137 (1999).
84. Taylor, R. *Straight Through from London: the Antipodes and Bounty Islands, New Zealand* (Heritage Expeditions New Zealand, 2006).
85. Universitat de les Illes Balears. *Herbario virtual del Mediterráneo Occidental* <http://herbarivirtual.uib.es/cas-med/index.html>
86. University of Kent. *Cook Islands Biodiversity and Ethnobiology Database* <http://cookislands.bishopmuseum.org/search.asp>
87. Wace, N. M. The vegetation of Gough Island. *Ecol. Monogr.* **31**, 337–367 (1961).
88. Wace, N. M. & Dickson, J. H. The terrestrial botany of the Tristan da Cunha Islands. *Phil. Trans. R. Soc. Lond. B* **249**, 273–360 (1965).
89. Wagner, W. L. & Lorence, D. H. *Flora of the Marquesas Islands* <http://botany.si.edu/pacificislandbiodiversity/marquesasflora/index.htm>
90. Wagner, W. L., Herbst, D. R. & Lorence, D. H. *Flora of the Hawaiian Islands* <http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/index.htm>
91. Whistler, W. A. *A Study of the Rare Plants of American Samoa* (US Fish and Wildlife Service, 1998).
92. Gerlach, J. The biodiversity of the granitic islands of Seychelles. *Phelsuma* **11** (Supplement A), 1–47 (2003).
93. Shaw, J. D., Spear, D., Greve, M. & Chown, S. L. Taxonomic homogenization and differentiation across Southern Ocean Islands differ among insects and vascular plants. *J. Biogeogr.* **37**, 217–228 (2010).
94. Dormann, C. F. et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609–628 (2007).
95. Crase, B., Liedloff, A. C. & Wintle, B. A. A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography* **35**, 879–888 (2012).
96. Bivand, R. *spdep: spatial dependence: weighting schemes, statistics and models v.0.5-77* (Package for R statistical software, 2014).
97. Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. Global diversity of island floras from a macroecological perspective. *Ecol. Lett.* **11**, 116–127 (2008).
98. Burnham, K. P. & Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*. 2nd edn (Springer, 2002).
99. Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol. Modell.* **135**, 147–186 (2000).
100. McFadden, D. in *Frontiers in Economics* (ed P. Zarembka) (Academic Press, 1974).
101. Efron, B. Regression and ANOVA with zero-one data: measures of residual variation. *J. Am. Stat. Assoc.* **73**, 113–121 (1978).

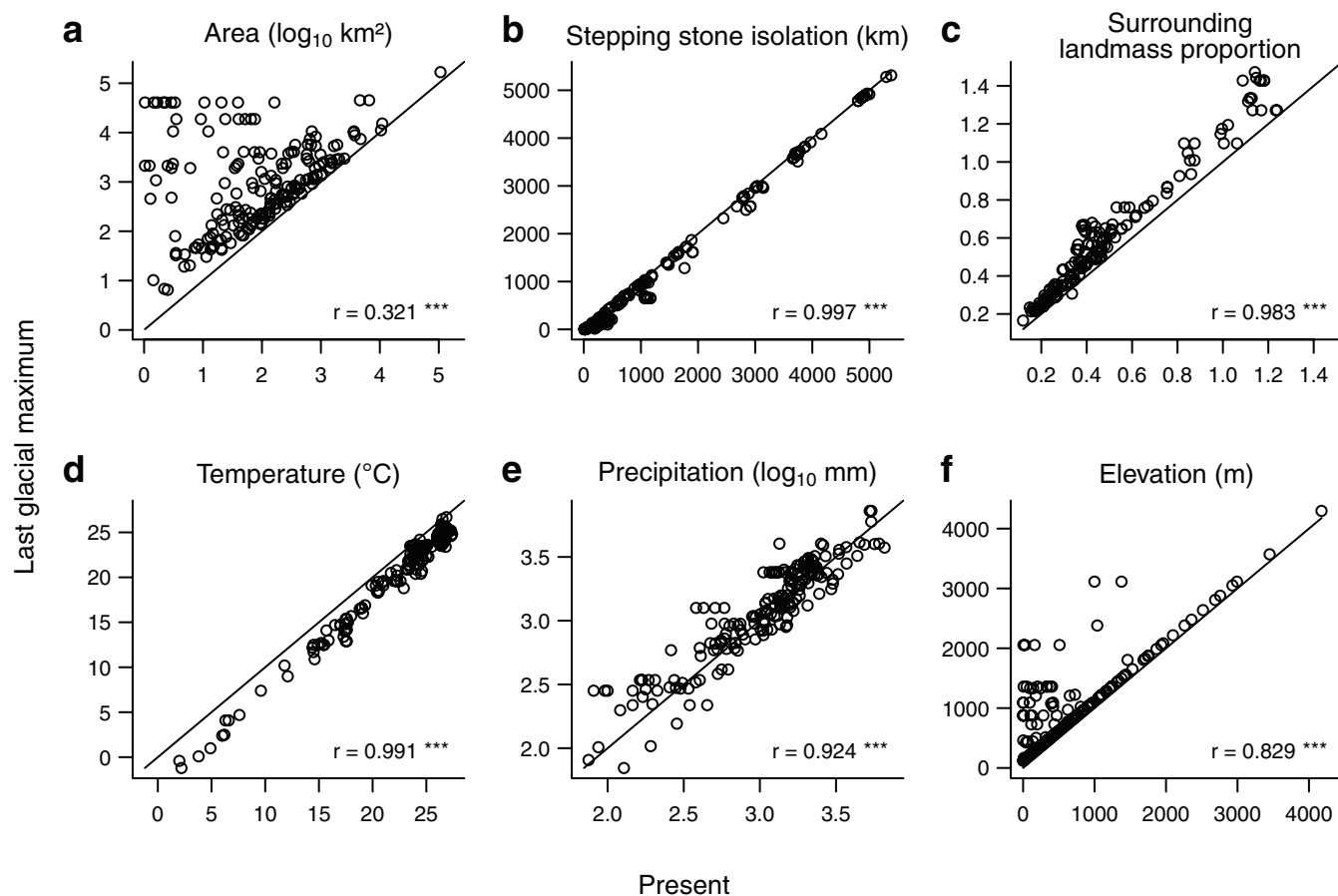


Extended Data Figure 1 | Frequency distributions of differences between last glacial maximum and present characteristics of 184 oceanic islands worldwide. a–e, Variables depicted by Δ give differences between island characteristics during the last glacial maximum and

today in island area (a), stepping-stone isolation (b), temperature (c), precipitation (d), and elevation (e). In f, no. of entities represents the number of present island entities that made up single past island units during the LGM owing to lower sea levels.

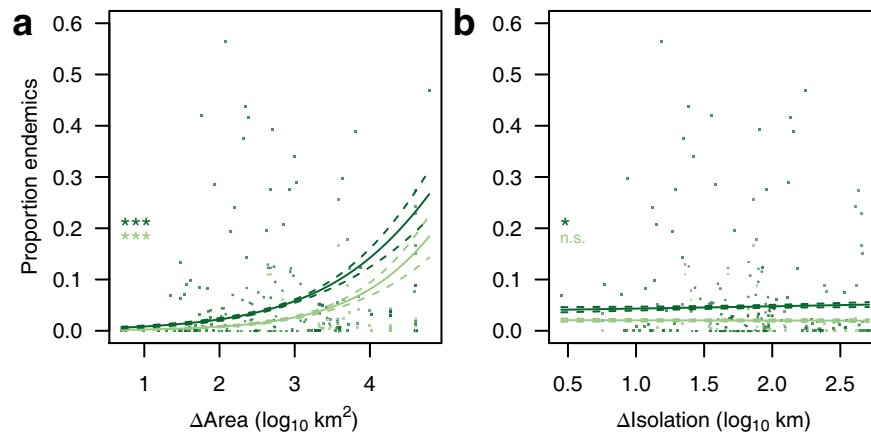


Extended Data Figure 2 | Geologic age and area of the four main geologic complexes of the Hawaiian Islands exemplifying the rate of area decrease with time due to subsidence and erosion for volcanic islands. In this example scenario, we assume that the four major complexes of the Hawaiian Islands (Fig. 2b) all reached approximately the same maximum area size and that the youngest complex, Big Island, already reached its full extent. The area decrease from Big Island to Maui Nui equals -0.0072 km^2 per year and from Maui Nui to O'ahu -0.0016 km^2 per year (see map in Fig. 2b). The linear fit over all points has a slope of -0.0016 km^2 per year. The negative exponential curve has a slope of -0.0034 km^2 per year at the beginning and of -0.0006 km^2 per year at the end. For comparison, the 184 islands used for this paper experienced an area decrease due to rising sea levels of 0.5035 km^2 per year over 10,000 years on average and the post-LGM decrease in island area for islands like Anegada or Mahé, was up to 1,000 times faster than the area decrease indicated here for the Hawaiian Islands. The notion that post-LGM changes have been much faster than average rates of geologic processes of island formation and erosion, therefore, most likely holds true even if the assumptions above are not perfectly met.



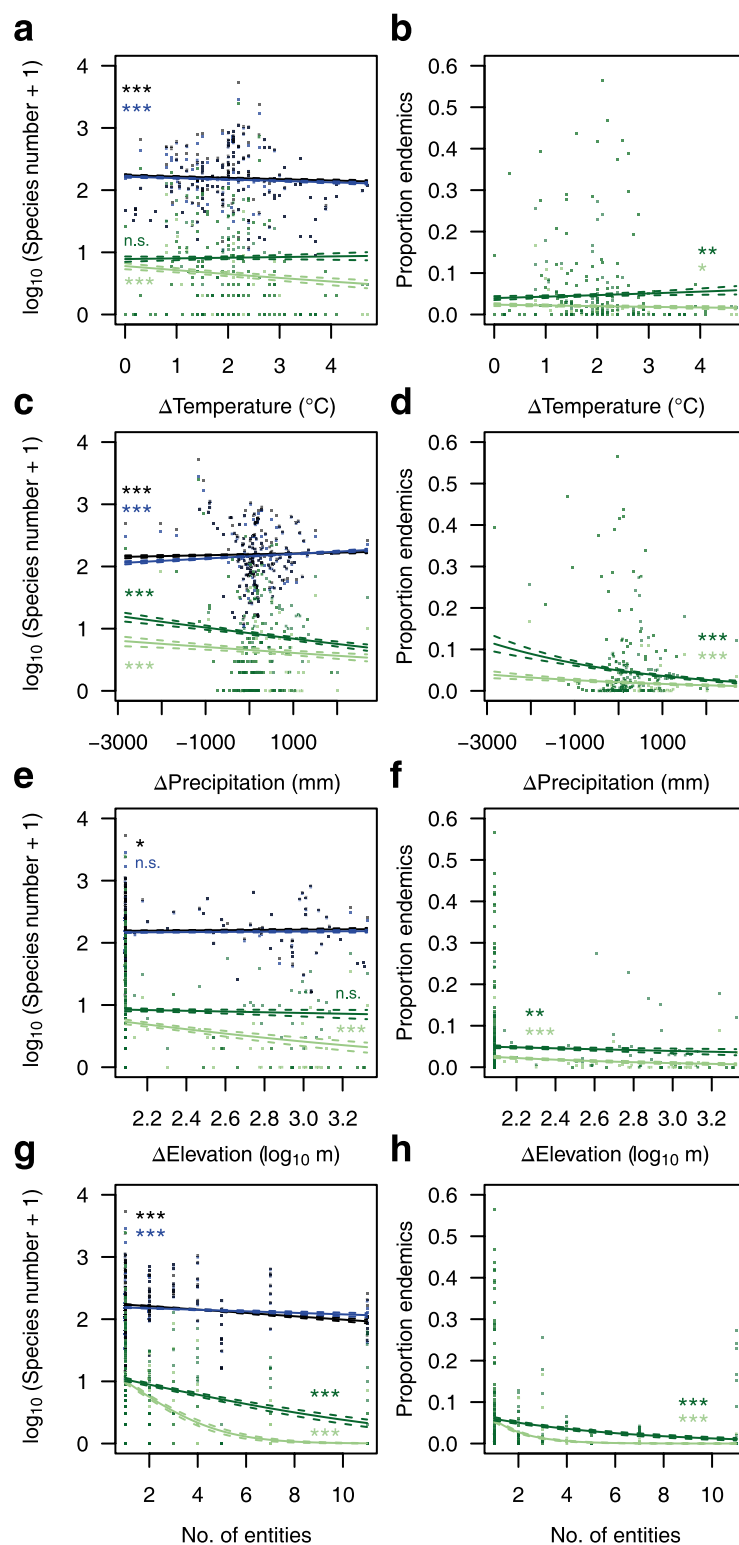
Extended Data Figure 3 | Pearson correlation coefficients (r) of last glacial maximum and present characteristics of 184 oceanic islands worldwide. a–f, Island area (a), stepping-stone isolation (b), isolation measured as proportion of surrounding landmass (c), temperature (d),

precipitation (e), and elevation (f). See Methods for variable descriptions. Diagonals indicate hypothetically equal last glacial maximum and present values. $***P < 0.001$.



Extended Data Figure 4 | Effects of post-LGM changes in biophysical island characteristics on the proportion of endemic species on 184 islands worldwide. a, b, Regression lines were predicted for changes in area (Δarea) (**a**) and stepping-stone isolation ($\Delta\text{isolation}$) (**b**) after accounting for all past and present covariables which were held constant at the mean of their empirical values across the islands (see Fig. 3 and Extended Data Fig. 5).

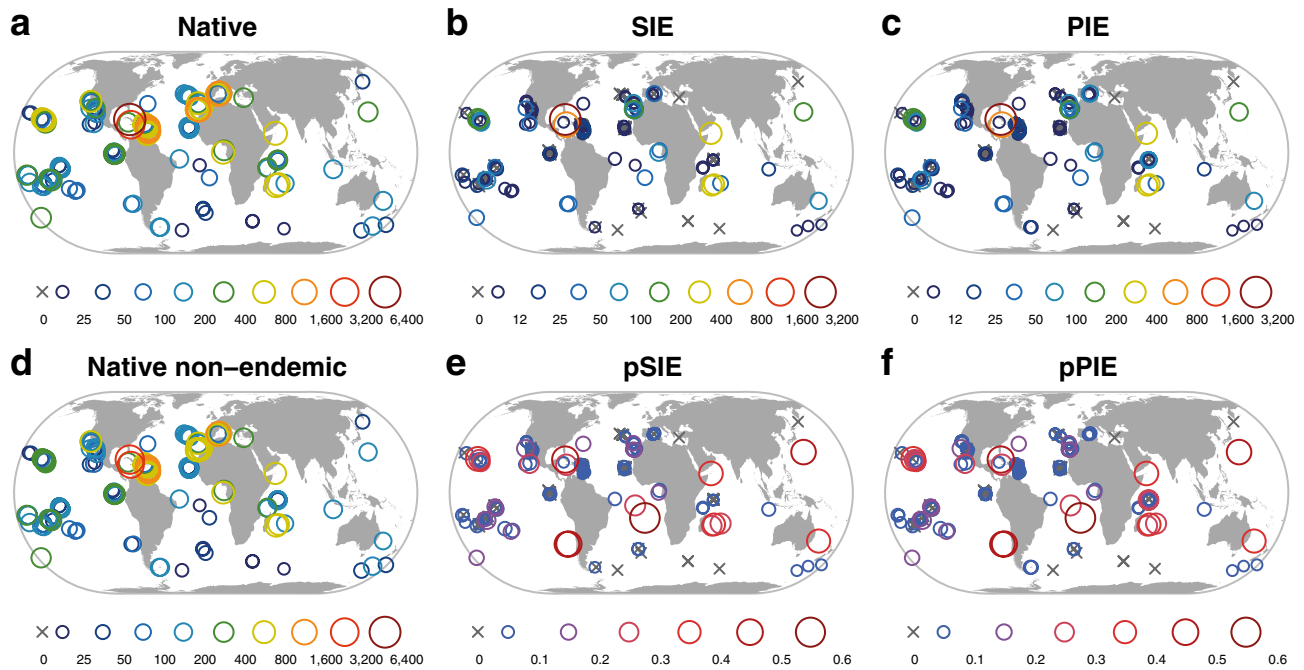
Relationships are shown for single-island endemics (light green) and species endemic to past island units (differs from single-island endemics in cases where several present islands originated from one island during the last glacial maximum) (green). Dashed lines indicate 95% confidence intervals.



Extended Data Figure 5 | Effects of post-LGM changes in biophysical island characteristics on angiosperm diversity on 184 islands worldwide.

Post-LGM changes (with Δ) give the difference in island characteristics from the LGM to today. No. of entities, number of present-island entities that made up single past island units during the LGM. Regression lines were predicted after accounting for all past and present covariables which were held constant at the mean of their empirical values across the islands (for Δ area and Δ isolation see Fig. 3 and Extended Data Fig. 4). In **a**, **c**, **e** and **g**, relationships are shown for

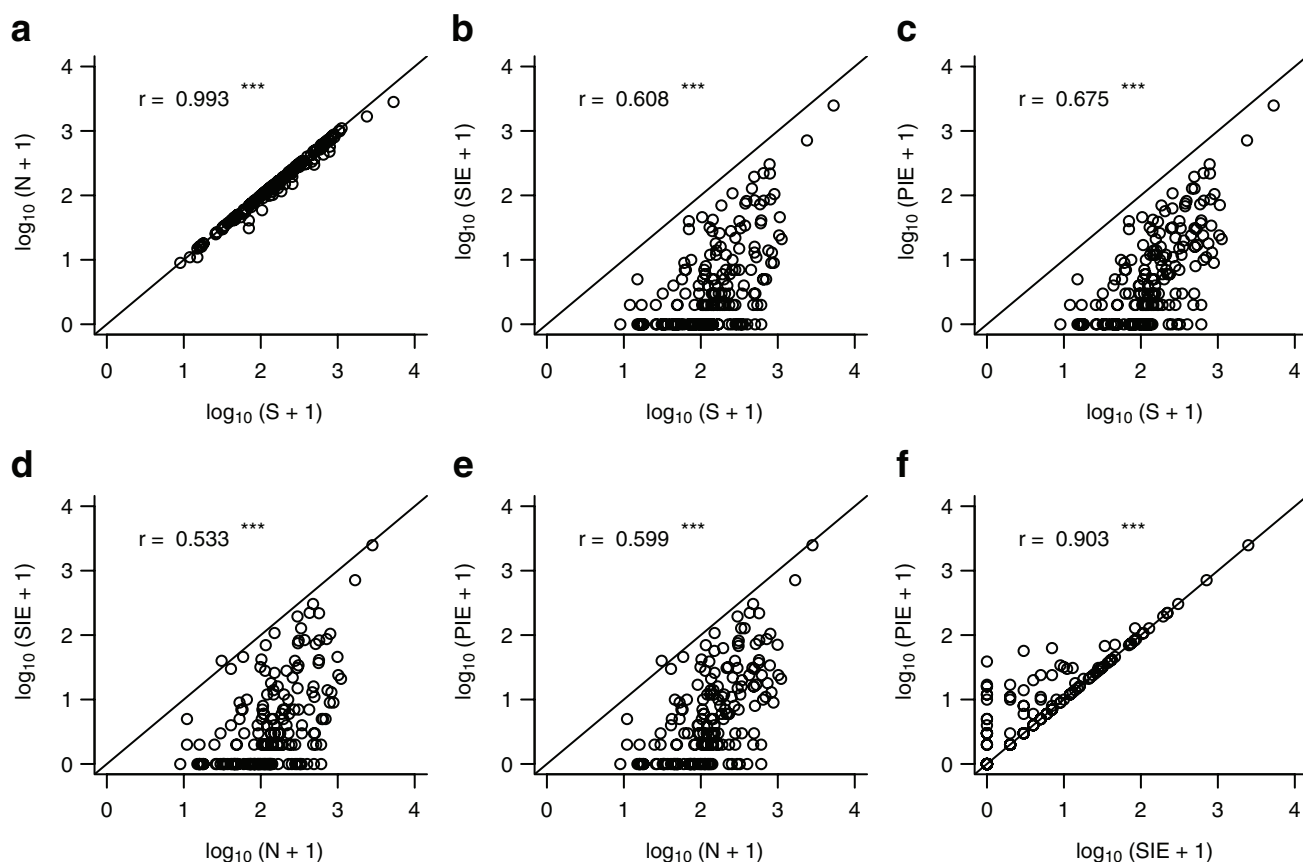
species numbers of natives (black), native-non-endemics (blue), single-island endemics (light green) and species endemic to past island units (green; differs from single-island endemics in cases where several present islands originated from one island during the LGM). In **b**, **d**, **f** and **h**, relationships are shown for proportions of single-island endemics (light green) and species endemic to past island units (green). Dashed lines indicate 95% confidence intervals. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, n.s., not significant ($P \geq 0.05$).



Extended Data Figure 6 | Species richness of native and endemic angiosperms on the 184 islands worldwide used for this study.

a–f, Native species (**a**), single-island endemics (SIE) (**b**), past-island-unit endemics (PIE; differs from SIE in cases where several present islands originated from one island during the last glacial maximum) (**c**), native non-endemics (**d**), proportion of single-island endemics (pSIE) (**e**), and

proportion of species endemic to past island units (pPIE) (**f**). Species richness is given in numbers of species. Numbers in legends indicate category borders. World maps are based on the GADM database of Global Administrative Areas, version 1 (<http://www.gadm.org/version1>). See Supplementary Data for values and see Methods for references used to compile the data set.



Extended Data Figure 7 | Relationships and Pearson correlation coefficients (r) of native and endemic species richness of angiosperms on 184 islands worldwide. a–c, Correlations of native species richness (S) with native non-endemics (N) (a), single-island endemics (SIE) (b) and past-island-unit endemics (PIE; differs from SIE in cases where several present islands originated from one island during the last glacial

maximum) (c). d–f, Correlations of native non-endemic species richness (N) with single-island endemics (SIE) (d) and past-island-unit endemics (PIE) (e) as well as richness of single-island endemics (SIE) with past-island-unit endemics (PIE) (f). Diagonals indicate lines of equal values. *** $P < 0.001$.

Extended Data Table 1 | Model statistics for best candidate models from AIC-based model selection for species richness of native (S), native non-endemic (N), single-island endemic (SIE), and past-island-unit endemic (PIE) angiosperm species and proportions of endemic species (pSIE, pPIE) in dependence on past and present island characteristics

S	Estimate	Std. error	z-value	P	pseudo-R²
Intercept	5.058	0.006	782.678	***	
log ₁₀ Area	0.413	0.010	41.389	***	D²_{adj}
Isolation	-0.521	0.007	-77.766	***	0.934
Temperature	0.344	0.007	51.157	***	
log ₁₀ Precipitation	0.320	0.006	52.234	***	McFadden
log ₁₀ Elevation	0.178	0.010	18.713	***	0.919
log ₁₀ ΔArea	0.262	0.011	24.164	***	
No. of entities	-0.167	0.011	-15.422	***	Efron
Spatial autocovariate	0.064	0.001	81.514	***	0.969
N					
Intercept	4.993	0.007	735.901	***	
log ₁₀ Area	0.421	0.007	57.525	***	D²_{adj}
Isolation	-0.576	0.007	-80.661	***	0.916
Temperature	0.306	0.007	42.926	***	
log ₁₀ Precipitation	0.331	0.007	47.765	***	McFadden
log ₁₀ Elevation	0.147	0.010	15.289	***	0.896
log ₁₀ ΔArea	0.085	0.007	12.520	***	
log ₁₀ ΔIsolation	-0.025	0.006	-4.153	***	Efron
Spatial autocovariate	0.068	0.001	81.542	***	0.950
SIE					
Intercept	1.372	0.051	27.114	***	
log ₁₀ Area	0.039	0.058	0.669		D²_{adj}
Isolation	-0.238	0.026	-9.158	***	0.921
Temperature	0.608	0.028	21.802	***	
log ₁₀ Precipitation	0.088	0.022	4.000	***	McFadden
log ₁₀ Elevation	1.045	0.061	17.216	***	0.913
log ₁₀ ΔArea	1.266	0.063	19.947	***	
log ₁₀ ΔIsolation	-0.059	0.019	-3.072	**	Efron
ΔTemperature	-0.164	0.028	-5.851	***	0.986
ΔPrecipitation	-0.088	0.019	-4.603	***	
log ₁₀ ΔElevation	-0.419	0.059	-7.058	***	
No. of entities	-1.821	0.102	-17.894	***	
Spatial autocovariate	0.229	0.007	33.276	***	
pSIE					
Intercept	-3.636	0.051	-71.385	***	
log ₁₀ Area	-0.499	0.065	-7.721	***	D²_{adj}
Temperature	0.241	0.033	7.286	***	0.830
log ₁₀ Precipitation	-0.265	0.025	-10.554	***	
log ₁₀ Elevation	1.205	0.068	17.670	***	McFadden
log ₁₀ ΔArea	0.736	0.061	11.974	***	0.809
ΔTemperature	-0.252	0.025	-10.023	***	
ΔPrecipitation	-0.223	0.019	-11.631	***	Efron
log ₁₀ ΔElevation	-0.254	0.060	-4.254	***	0.596
No. of entities	-1.766	0.103	-17.165	***	
Spatial autocovariate	0.233	0.006	39.781	***	
PIE					
Intercept	2.190	0.029	74.377	***	
log ₁₀ Area	0.260	0.045	5.834	***	D²_{adj}
Isolation	-0.068	0.025	-2.746	**	0.918
Temperature	0.576	0.024	24.225	***	
log ₁₀ Precipitation	0.062	0.021	2.979	**	McFadden
log ₁₀ Elevation	0.762	0.047	16.244	***	0.906
log ₁₀ ΔArea	1.080	0.048	22.715	***	
log ₁₀ ΔIsolation	-0.034	0.017	-1.969	*	Efron
ΔTemperature	0.012	0.024	0.497		0.986
ΔPrecipitation	-0.201	0.015	-13.144	***	
No. of entities	-0.542	0.038	-14.313	***	
Spatial autocovariate	0.230	0.005	46.127	***	
pPIE					
Intercept	-2.872	0.031	-91.579	***	
log ₁₀ Area	-0.177	0.054	-3.265	**	D²
Temperature	0.319	0.027	11.829	***	0.805
log ₁₀ Precipitation	-0.171	0.023	-7.341	***	
log ₁₀ Elevation	0.682	0.050	13.521	***	McFadden
log ₁₀ ΔArea	0.732	0.059	12.353	***	0.777
log ₁₀ ΔIsolation	0.193	0.021	9.071	***	
ΔTemperature	-0.165	0.024	-6.885	***	Efron
ΔPrecipitation	-0.188	0.018	-10.562	***	0.520
log ₁₀ ΔElevation	-0.031	0.036	-0.860		
No. of entities	-0.512	0.045	-11.388	***	
Spatial autocovariate	0.305	0.007	44.398	***	

Variables that quantify the change in island characteristics between the LGM and present are depicted by Δ. No. of entities represents the number of present island entities that made up single past island units during the LGM due to lower sea levels. Estimates are based on zero-mean-unit-variance standardized predictor variables for comparison (except for the spatial autocovariate). All best models were characterized by AIC values more than 7 units smaller than those for the second-best models. As measures of variability explained we provide pseudo- R^2 values: D^2_{adj} , deviance explained adjusted for sample size and number of predictors⁹⁹, McFadden, McFadden's pseudo R^2 (ref. 100), Efron, R^2 -value of a linear model between the observed and predicted values of the generalized linear models (GLM)¹⁰¹. Please note that Efron's pseudo- R^2 values for Poisson GLMs (S, N, SIE, PIE) probably overestimate the variability explained as the linear models between predicted and observed values were mainly driven by few large values. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Extended Data Table 2 | Matrix of Pearson correlation coefficients of all 11 past and present environmental predictors used in this study ($n = 184$ islands)

	\log_{10} Area	Isolation	Temperature	\log_{10} Precipitation	\log_{10} Elevation	$\log_{10} \Delta$ Area	$\log_{10} \Delta$ Isolation	Δ Temperature	Δ Precipitation	$\log_{10} \Delta$ Elevation
Isolation	-0.060									
Temperature	-0.121	0.127 #								
\log_{10} Precipitation	0.178 *	0.420 ***	0.128 #							
\log_{10} Elevation	0.735 ***	0.062	-0.207 **	0.218 **						
$\log_{10} \Delta$ Area	0.188 *	-0.334 ***	0.056	0.055	-0.074					
$\log_{10} \Delta$ Isolation	0.008	-0.036	0.013	0.482 ***	-0.085	0.461 ***				
Δ Temperature	-0.004	-0.364 ***	-0.439 ***	-0.131 #	0.141 #	-0.049	-0.003			
Δ Precipitation	-0.353 ***	0.220 **	0.149 *	-0.043	-0.255 ***	0.060	0.004	-0.205 **		
$\log_{10} \Delta$ Elevation	-0.371 ***	-0.179 *	0.145 #	-0.133 #	-0.378 ***	0.602 ***	0.200 **	-0.127 #	0.387 ***	
No. of entities	-0.350 ***	-0.208 **	0.206 **	0.032	-0.374 ***	0.715 ***	0.399 ***	-0.043	0.297 ***	0.704 ***

Past variables are depicted by Δ and give the difference between island characteristics during the last glacial maximum and today. No. of entities represents the number of present island entities that made up single past island units during the LGM due to lower sea levels. # $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

A specific area of olfactory cortex involved in stress hormone responses to predator odours

Kunio Kondoh^{1*}, Zhonghua Lu^{1*}, Xiaolan Ye¹, David P. Olson^{2†}, Bradford B. Lowell² & Linda B. Buck¹

Instinctive reactions to danger are critical to the perpetuation of species and are observed throughout the animal kingdom. The scent of predators induces an instinctive fear response in mice that includes behavioural changes, as well as a surge in blood stress hormones that mobilizes multiple body systems to escape impending danger^{1,2}. How the olfactory system routes predator signals detected in the nose to achieve these effects is unknown. Here we identify a specific area of the olfactory cortex in mice that induces stress hormone responses to volatile predator odours. Using monosynaptic and polysynaptic viral tracers, we found that multiple olfactory cortical areas transmit signals to hypothalamic corticotropin-releasing hormone (CRH) neurons, which control stress hormone levels. However, only one minor cortical area, the amygdalo-piriform transition area (AmPir), contained neurons upstream of CRH neurons that were activated by volatile predator odours. Chemogenetic stimulation of AmPir activated CRH neurons and induced an increase in blood stress hormones, mimicking an instinctive fear response. Moreover, chemogenetic silencing of AmPir markedly reduced the stress hormone response to predator odours without affecting a fear behaviour. These findings suggest that AmPir, a small area comprising <5% of the olfactory cortex, plays a key part in the hormonal component of the instinctive fear response to volatile predator scents.

Animals sense environmental dangers that challenge their survival and respond with behavioural and physiological changes to eliminate the challenge. The olfactory system detects chemical cues emitted from predators that signal danger and stimulate an instinctive fear response. In addition to characteristic behaviours, this response includes increases in blood stress hormones that are reminiscent of responses to fear and stress in humans and are controlled by CRH neurons in the paraventricular nucleus of the hypothalamus (PVN)^{1–3}. The neural circuits that convey predator odour signals from the nose to CRH neurons are presently unknown.

Volatile predator cues are detected in the olfactory epithelium of the nose, which also detects common odorants, whereas fear-stimulating proteins are detected in the vomeronasal organ (VNO), an accessory olfactory structure^{4,5}. Sensory signals travel from the nose through the olfactory bulb to the olfactory cortex (OC) and from the VNO through the accessory olfactory bulb to the vomeronasal amygdala (VA)^{6,7}. The mouse nose has ~1,000 different odorant receptors, each expressed by a unique subset of dispersed sensory neurons^{8–11}. The olfactory bulb contains a semi-stereotyped map of receptor inputs in which individual glomeruli and projection neurons appear to be dedicated to one receptor^{12–14}, but single bulb neurons can project widely in the OC^{15–17}. The OC comprises multiple distinct areas whose respective functions are largely obscure^{6,7}. Recent studies have linked an amygdala OC area to odour attraction and aversion¹⁸. We investigated whether a specific OC area governs stress hormone responses to predator odours.

We first asked which OC areas can transmit signals to CRH neurons. We developed two Cre-recombinase-dependent Bartha pseudorabies viruses (PRVs) that have nonfunctional thymidine kinase (TK) genes, but require TK to replicate and travel retrogradely across synapses^{19,20}, and a lentivirus (LVF2TK) with Cre-dependent expression of haemagglutinin-tagged TK (HA-TK) (Fig. 1a). PRVB177 is a ‘polysynaptic PRV’ with irreversible Cre-dependent expression of HA-TK that can cross multiple sequential synapses. PRVB316 is a ‘monosynaptic PRV’ with Cre-dependent expression of GFP that can replicate in neurons co-infected with LVF2TK and spread across one synapse, but not further. We injected the viruses into the PVN of mice that express Cre in CRH neurons (CRH-IRES-Cre (CRH-Cre) mice)²¹ (Fig. 1b) and then immunostained brain sections for HA (PRVB177) or GFP (PRVB316) to visualize virus-infected neurons.

Preliminary experiments confirmed that Cre is expressed in PVN CRH neurons in CRH-Cre mice and that the viruses function as expected. In CRH-Cre mice crossed with Rosa-floxed-GFP mice, 93.9 ± 0.5% of GFP⁺ cells expressed *Crh* (*n* = 4) (Extended Data Fig. 1a). After PVN co-injection of PRVB316 and LVF2TK, some neurons expressed both viruses (Extended Data Fig. 1b). Neurons infected with PRVB316, but not LVF2TK, were subsequently detected outside the PVN, whereas none were detected outside the PVN when PRVB316 was injected without LVF2TK (Extended Data Fig. 1c). No infected neurons were seen in wild-type animals injected with the monosynaptic or polysynaptic PRV (Extended Data Fig. 1c, d).

Virus-infected neurons were detected in multiple brain areas after PRV infection of CRH neurons (Extended Data Figs 2 and 3). The monosynaptic PRV (PRVB316) infected 31 brain areas, including areas known to project axons to the PVN²² and areas linked to CRH neuron activation by certain non-olfactory stressors³.

In the olfactory system (Fig. 1c), PRVB316 infected only the posterior medial amygdala (MEAp), part of the VA (Fig. 1d, e), whereas, by day 4 after injection, PRVB177 infected five OC and both VA areas (Fig. 1d, f, g). Most brain areas infected with PRVB177 on day 3 were also infected by PRVB316 (Extended Data Figs 2 and 3), suggesting that OC areas infected on day 4 may transmit signals to CRH neurons via one intermediate relay.

We observed a regional bias in the locations of neurons upstream of CRH neurons in some olfactory areas, but not others. PRVB177⁺ neurons were randomly distributed along the anterior–posterior axis in most olfactory areas, but were concentrated in posterior regions of the posterior piriform cortex (pPir) and AmPir and in the posteroven-tral quadrant of the MEA (Extended Data Fig. 4).

Dual labelling for PRVB177 and markers of excitatory glutamatergic neurons (*Vglut1* (also known as *Slc17a7*) and *Vglut2* (also known as *Slc17a6*)) or inhibitory GABAergic neurons (*Gad1* and *Gad2*) indicated that some neurons upstream of CRH neurons in the MEA were GABAergic, but the majority of those labelled for the markers in higher olfactory areas were glutamatergic (Extended Data Fig. 5). The

¹Howard Hughes Medical Institute, Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. ²Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02115, USA. [†]Present address: Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, Michigan 48109, USA.

*These authors contributed equally to the work.

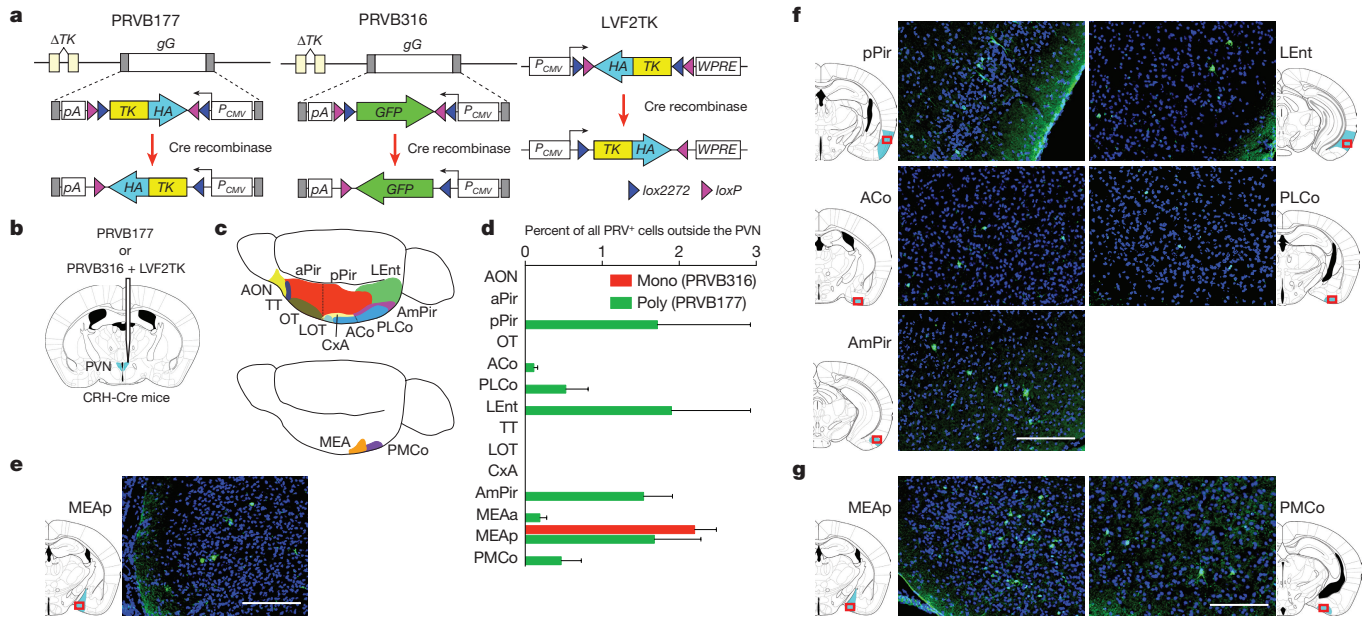


Figure 1 | Higher olfactory areas contain neurons upstream of CRH neurons. **a**, Cre recombinase causes irreversible expression of HA-TK from PRVB177 and LVF2TK, and GFP from PRVB316. gG, gG locus; P_{CMV} , cytomegalovirus promoter; pA, polyadenylation signal; WPRE, woodchuck hepatitis post-transcriptional regulatory element. **b**, Viruses were injected into the PVN (blue) of CRH-Cre mice. **c**, The OC (top) comprises multiple

areas and the VA (bottom) two areas. See Methods for full names. **d**, Mean percentages of PRV⁺ neurons in olfactory areas 4 days after injection of PRVB316 ($n = 24$) or PRVB177 ($n = 6$). Error bars indicate standard error of the mean (s.e.m.). **e–g**, Photographs and diagrams of olfactory area (blue) sections³⁰ (red rectangles) with neurons immunostained for PRVB316 (**e**) or PRVB177 (**f, g**) (green). Scale bars, 200 μ m.

glutamatergic neurons could potentially activate either excitatory or inhibitory neurons directly presynaptic to CRH neurons in other brain areas, and thus either stimulate or inhibit CRH neurons.

Neurons upstream of CRH neurons in different OC areas could conceivably transmit redundant predator odour signals to CRH neurons. To investigate this possibility, we used a neural activity marker, nuclear *Arc* messenger RNA (*nArc*)²³ to identify PRVB177-infected neurons activated by the fox predator odourant 2,5-dihydro-2,4,5-trimethylthiazoline (TMT)^{4,24} or bobcat (predator) urine. Preliminary experiments confirmed that TMT and bobcat urine, but not non-predator (rabbit) urine, induce the activity marker *c-Fos* in PVN CRH neurons and increase blood levels of the stress hormones adrenocorticotrophic hormone (ACTH) and corticosterone (Extended Data Fig. 6).

Surprisingly, both predator odours induced a significant increase in *nArc* in PRV⁺ neurons in only one OC area, AmPir (Fig. 2). The percentage of AmPir PRV⁺ neurons expressing *nArc* was increased 5.8-fold by TMT and 4.9-fold by bobcat urine, whereas rabbit urine had no significant effect. These results suggested that AmPir could play an important part in hormonal fear responses to volatile predator odours.

To explore this idea, we first asked whether chemogenetic activation of AmPir induces stress hormone increases. We injected the adeno-associated virus (AAV), AAV-DIO-hM3Dq-mCherry²⁵ bilaterally into the AmPir, pPir or MEA of mice expressing Cre in projection neurons (Emx1-IRES-Cre or Vglut2-IRES-Cre mice)^{26,27} (Fig. 3a–d). This virus has Cre-dependent expression of mCherry fused to hM3Dq, a receptor that depolarizes neurons upon binding its pharmacologically inert ligand, clozapine-*N*-oxide (CNO)²⁸. We then injected animals with CNO or saline and measured plasma ACTH.

Chemogenetic activation of AmPir induced a striking (7.6-fold) increase in plasma ACTH (Fig. 3e). Activation of MEA, but not pPir, also increased plasma ACTH. Consistent with these results, activation of AmPir or MEA also induced *c-Fos* in PVN CRH neurons (Fig. 3f, g). Additional analyses showed that mCherry⁺ (AAV-infected) neurons were present in a large part of each targeted area (Extended Data Fig. 7) and that CNO induced *nArc* in neurons in each injected area (Fig. 3h, i). Those in the AmPir could include neurons responsive to TMT or bobcat urine as well as neurons activated by other odorants. Owing to the

relatively small size of the AmPir, infected neurons were also often seen in immediately adjacent olfactory areas (parts of the posterior posterolateral cortical amygdala (PLCo) and lateral posteromedial cortical amygdala (PMCo)). These results indicated that activation of neurons in and around AmPir induces a stress hormone response that mimics the instinctive fear response.

We next used chemogenetic silencing to ask whether the AmPir is required for stress hormone responses to predator odours. We injected

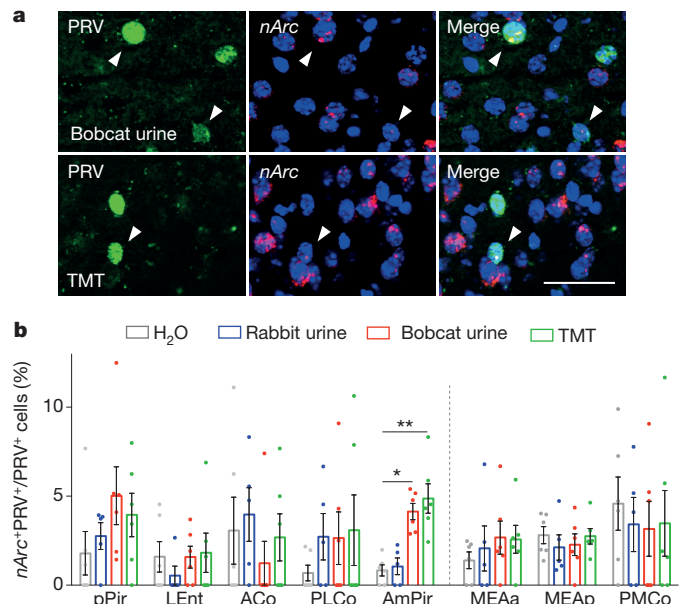


Figure 2 | Predator odours activate neurons upstream of CRH neurons in AmPir. **a**, AmPir neurons co-labelled for PRVB177 (green) and *nArc* (red) after exposure of mice to bobcat urine or TMT. Scale bar, 50 μ m. **b**, Percentage of PRVB177⁺ neurons co-labelled for *nArc* in areas of the OC (left) and VA (right) after exposure to TMT, bobcat urine, rabbit urine or water. $n = 5–6$ per condition. Error bars indicate s.e.m. * $P < 0.05$, ** $P < 0.01$, Kruskal–Wallis test with post-hoc Dunn's test.

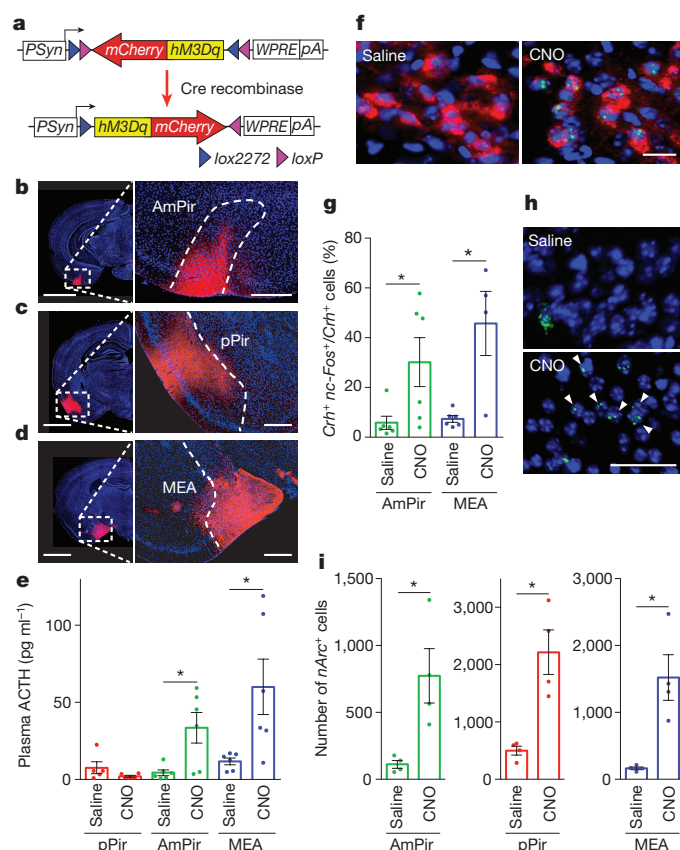


Figure 3 | Chemogenetic activation of the AmPir induces stress hormone increase. **a–d**, mCherry immunostaining (red) after injection of AAV-DIO-hM3Dq-mCherry (**a**) into the AmPir (**b**), pPir (**c**) or MEA (**d**). *PSyn*, human synapsin 1 promoter. Scale bars, 2 mm (left), 400 μ m (right). **e**, Plasma ACTH concentration after injection of CNO or saline into mice expressing hM3Dq in the pPir, AmPir or MEA. $n = 5–6$ per condition. Error bars indicate s.e.m. * $P < 0.05$, Mann–Whitney *U*-test. **f**, PVN neurons co-labelled for *Crh* (red) and nuclear *c-Fos* (*nc-Fos*) (green) after injection of saline or CNO into mice expressing hM3Dq in the AmPir. Scale bar, 20 μ m. **g**, Percentage of PVN *Crh*⁺ neurons co-labelled for *nc-Fos* in mice treated as in **e**. $n = 4–6$ per condition. Error bars indicate s.e.m. * $P < 0.05$, Mann–Whitney *U*-test. **h**, AmPir neurons labelled for *nAra* (green) (arrowheads) after saline or CNO injection of mice expressing hM3Dq in the AmPir. Scale bar, 100 μ m. **i**, Number of *nAra*⁺ neurons in the AmPir, pPir or MEA in mice treated as in **e**. $n = 4$ per condition. Error bars indicate s.e.m. * $P < 0.05$, unpaired *t*-test.

the AmPir of *Emx1*-Cre mice bilaterally with AAV, AAV-DIO-hM4Di-mCherry²⁵ (Fig. 4a, b and Extended Data Fig. 8a, b). This virus has Cre-dependent expression of hM4Di, which silences neurons upon binding to CNO²⁸. We subsequently injected mice with CNO or saline, exposed them to predator odours, and then determined plasma levels of stress hormones.

Silencing of the AmPir dramatically reduced stress hormone responses to predator odours (Fig. 4d). Plasma ACTH responses to TMT and bobcat urine were decreased by $52 \pm 11\%$ and $55 \pm 02\%$, respectively (Fig. 4d). These decreases correlated with reductions of $58 \pm 03\%$ and $65 \pm 04\%$ in the percentages of AmPir neurons activated by TMT and bobcat urine, as assessed by *nAra* labelling (Extended Data Fig. 9a, b). AmPir silencing also decreased plasma corticosterone (Fig. 4e) and the percentage of *c-Fos*⁺ PVN CRH neurons observed after predator odour exposure (Fig. 4f). These results suggested that neurons in the AmPir are required for full predator-odour-induced increases in stress hormones.

To examine whether adjacent PLCo/PMCo regions were involved in the silencing effects observed for the AmPir, we injected AAV-DIO-hM4Di-mCherry selectively into these regions, but not into

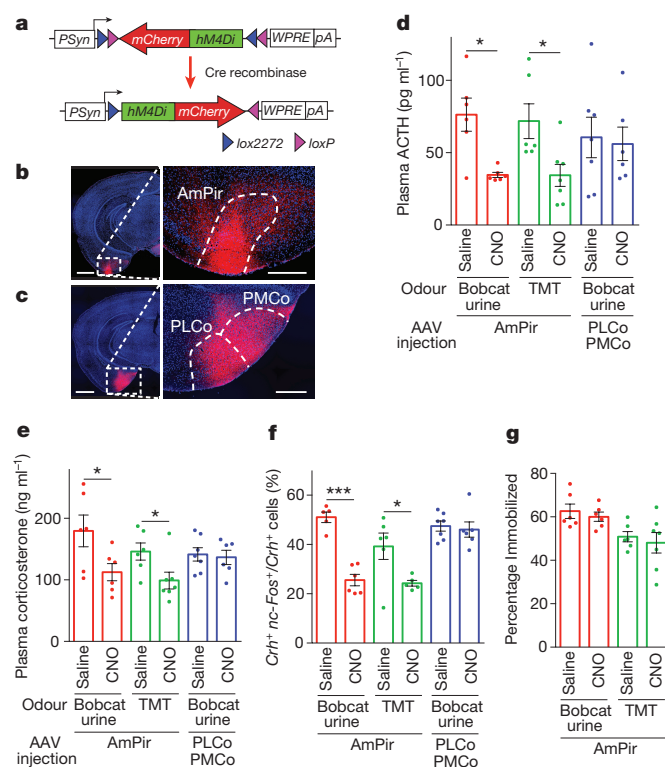


Figure 4 | Chemogenetic silencing of the AmPir inhibits stress hormone responses to predator odours. **a–c**, mCherry immunostaining (red) after injection of AAV-DIO-hM4Di-mCherry (**a**) into the AmPir (**b**) or PLCo/PMCo (**c**). Scale bars, 1 mm (left), 400 μ m (right). **d**, **e**, Plasma ACTH (**d**) or corticosterone (**e**) concentration after mice expressing hM4Di in the AmPir or PLCo/PMCo were injected with saline or CNO and then exposed to bobcat urine or TMT. $n = 6–7$ per condition. Error bars indicate s.e.m. * $P < 0.05$, Mann–Whitney *U*-test (**d**) or unpaired *t*-test (**e**). **f**, Percentage of PVN *Crh*⁺ neurons expressing *nc-Fos* in mice treated as in **d**, $n = 5–7$ per condition. Error bars indicate s.e.m. * $P < 0.05$, *** $P < 0.001$, unpaired *t*-test. **g**, Mice expressing hM4Di in the AmPir (**d–f**) were videotaped during exposure to bobcat urine and videos were analysed for the percentage of time spent immobile (percentage immobilized). $n = 6–7$ per condition. Error bars indicate s.e.m.

the AmPir (Fig. 4c and Extended Data Fig. 8c). In sharp contrast to silencing of the AmPir, silencing of these regions did not affect predator-odour-induced increases in ACTH and corticosterone or CRH neuron activation (Fig. 4d–f and Extended Data Fig. 9c). These results clearly indicate that the AmPir, but not adjacent regions of the PLCo or PMCo, is involved in predator-odour-induced hormonal fear responses.

Interestingly, AmPir silencing did not affect the ‘freezing’ behaviour (immobilization) induced by bobcat urine or TMT, a typical fear behaviour caused by predator odours (Fig. 4g). A recent study showed that inhibition of neurons in the PLCo reduced freezing behaviour to TMT¹⁸. Given the effects of AmPir silencing on stress hormone responses to predator odours, these findings suggest that different OC areas may be involved in the behavioural versus hormonal effects of predator odours, with the AmPir and PLCo having major roles in their hormonal and behavioural effects, respectively.

Instinctive fear responses to predator odours are observed in naive animals, suggesting the involvement of stereotyped olfactory circuits. We have used a combination of monosynaptic and polysynaptic viral tracing, indicators of neuronal activation, and chemogenetic activation and silencing to investigate how the OC translates volatile predator odour signals from the nose into a hormonal fear response. These studies implicate only one small area of the OC in this response, the AmPir. They reveal that neurons in multiple OC areas can transmit signals to CRH neurons that control stress hormones, but that

fox and bobcat predator odours significantly activate only those in the AmPir. Chemogenetic activation of the AmPir induces stress hormone increases and its silencing causes a striking decrease in the stress hormone response to predator odours. Together, these findings suggest that the AmPir plays a key part in generating hormonal fear responses to volatile predator cues that signal danger. The functions of neurons upstream of CRH neurons in other OC areas are unknown. They might also contribute to stress hormone responses. However, given that rose oil can block stress hormone responses to a predator odour²⁹, it is also conceivable that some transmit signals that suppress rather than activate hormonal responses associated with fear.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 June 2015; accepted 20 January 2016.

Published online 21 March 2016.

1. Apfelbach, R., Blanchard, C. D., Blanchard, R. J., Hayes, R. A. & McGregor, I. S. The effects of predator odors in mammalian prey species: a review of field and laboratory studies. *Neurosci. Biobehav. Rev.* **29**, 1123–1144 (2005).
2. Takahashi, L. K. Olfactory systems and neural circuits that modulate predator odor fear. *Front. Behav. Neurosci.* **8**, 72 (2014).
3. Ulrich-Lai, Y. M. & Herman, J. P. Neural regulation of endocrine and autonomic stress responses. *Nature Rev. Neurosci.* **10**, 397–409 (2009).
4. Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450**, 503–508 (2007).
5. Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141**, 692–703 (2010).
6. Buck, L. B. & Bargmann, C. in *Principles of Neuroscience* (eds Kandel, E., Schwartz, J., Jessell, T., Siegelbaum, S. & Hudspeth, A. J.) 712–742 (McGraw-Hill, 2012).
7. Neville, K. R. & Haberly, L. B. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 415–454 (Oxford University Press, 2004).
8. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
9. Ressler, K. J., Sullivan, S. L. & Buck, L. B. A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597–609 (1993).
10. Vassar, R., Ngai, J. & Axel, R. Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* **74**, 309–318 (1993).
11. Hanchate, N. K. *et al.* Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science* **350**, 1251–1255 (2015).
12. Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1255 (1994).
13. Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981–991 (1994).
14. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
15. Sosulski, D. L., Bloom, M. L., Cutforth, T., Axel, R. & Datta, S. R. Distinct representations of olfactory information in different cortical centres. *Nature* **472**, 213–216 (2011).
16. Miyamichi, K. *et al.* Cortical representations of olfactory input by trans-synaptic tracing. *Nature* **472**, 191–196 (2011).
17. Ghosh, S. *et al.* Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature* **472**, 217–220 (2011).
18. Root, C. M., Denny, C. A., Hen, R. & Axel, R. The participation of cortical amygdala in innate, odour-driven behaviour. *Nature* **515**, 269–273 (2014).
19. Card, J. P. & Enquist, L. W. Transneuronal circuit analysis with pseudorabies viruses. *Curr. Protoc. Neurosci.* Chapter 1, Unit 1.5 (2001).
20. DeFalco, J. *et al.* Virus-assisted mapping of neural inputs to a feeding center in the hypothalamus. *Science* **291**, 2608–2613 (2001).
21. Krashes, M. J. *et al.* An excitatory paraventricular nucleus to AgRP neuron circuit that drives hunger. *Nature* **507**, 238–242 (2014).
22. Sawchenko, P. E. & Swanson, L. W. The organization of forebrain afferents to the paraventricular and supraoptic nuclei of the rat. *J. Comp. Neurol.* **218**, 121–144 (1983).
23. Guzowski, J. F., McNaughton, B. L., Barnes, C. A. & Worley, P. F. Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. *Nature Neurosci.* **2**, 1120–1124 (1999).
24. Vernet-Maury, E. in *Olfaction and Taste* (ed. Van der Starre, H.) 407 (IRL, 1980).
25. Krashes, M. J. *et al.* Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *J. Clin. Invest.* **121**, 1424–1428 (2011).
26. Gorski, J. A. *et al.* Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).
27. Vong, L. *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71**, 142–154 (2011).
28. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potentially activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104**, 5163–5168 (2007).
29. Matsukawa, M., Imada, M., Murakami, T., Aizawa, S. & Sato, T. Rose odor can innately counteract predator odor. *Brain Res.* **1381**, 117–123 (2011).
30. Franklin, K. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd edn (Academic, 2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Enquist for providing the parental PRV, PRV TK-BaBlu, and for helpful advice, and members of the Buck laboratory for discussions and comments. This work was supported by the Howard Hughes Medical Institute (HHMI) and by grants from the National Institutes of Health (National Institute on Deafness and Other Communication Disorders) (L.B.B.) and the Japan Society for the Promotion of Science (K.K.). L.B.B. is an Investigator of the HHMI.

Author Contributions K.K. and L.B.B. conceived the project, Z.L. and K.K. designed and prepared the viruses, X.Y. performed experiments, D.P.O. and B.B.L. generated the CRH-Cre mice, K.K. developed methods and performed most of the experiments and data analysis, and K.K. and L.B.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.B.B. (lbuck@fhcrc.org).

METHODS

Mice. CRH-IRES-Cre mice were generated previously²¹. C57BL/6J wild-type mice, Emx1-IRES-Cre and Vglut2-IRES-Cre knock-in mice, and Rosa-floxed-GFP reporter mice were purchased from the Jackson Laboratory. All procedures involving mice were approved by the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee.

We used both males and females in all experiments, with similar numbers where possible. The same number of each sex was used in experiments shown in Fig. 1 and Extended Data Figs 2, 3, 5 and 6.

No statistical methods were used to predetermine sample size. Animals were randomly chosen for experimental subjects. Animals were excluded from certain experiments. For *nArc* experiments shown in Fig. 2, animals were excluded if PRV⁺ cells were not found in all 5 OC and 3 VA areas analysed. For chemogenetic activation or silencing of the AmPir, animals were excluded if more than approximately 50% of mCherry⁺ cells were observed outside the AmPir.

Viral vectors. PRVs. PRVB177 and PRVB316 were constructed using homologous recombination between targeting vectors and genomic DNA of PRV TK-BaBlu, a thymidine kinase (TK)-deleted PRV Bartha strain derivative with a LacZ insertion into the *gG* locus²⁰. For targeting vectors, a flexstop-flanked sequence³¹ encoding a PRV TK fused at its C terminus to a haemagglutinin (HA) epitope tag (for PRVB177), or enhanced green fluorescent protein (eGFP) (for PRVB316), was first cloned with an inverse orientation into an eGFP-deleted pEGFP-N1 vector (Clontech). Next, *NsiI* fragments containing a CMV promoter, the flexstop-flanked coding sequence, and an SV40 polyadenylation signal were cloned between *gG* locus sequences matching those 5' and 3' to the lacZ sequence in PRV TK-BaBlu to give the final targeting vectors. These vectors were then linearized and co-transfected with PRV TK-BaBlu genomic DNA into HEK 293T cells (ATCC). Recombinant virus clones were selected and confirmed following methods described previously³².

To propagate recombinant PRVs, PK15 cells (ATCC) were infected with the viruses using a multiplicity of infection (m.o.i.) = 0.1–0.01. After being infected, cells showed a prominent cytopathic effect (~2 days). They were harvested by scraping, and the cell material was frozen using liquid nitrogen and then quickly thawed in a 37°C water bath. After three freeze–thaw cycles, cell debris was removed by centrifugation twice at 1,000g for 5 min and the supernatant was then used for experiments. The titre of viral stocks was determined using standard plaque assays on PK15 cells¹⁹, with titres expressed in plaque-forming units (p.f.u.). **Lentivirus.** To generate LVF2TK, a flexstop-flanked sequence encoding TK–HA (see earlier) was cloned into the pLenti6.3 vector (Thermo Fisher). LVF2TK was produced using the ViraPower HiPerform Lentiviral Expression System (Thermo Fisher) according to the manufacturer's instructions. Virus was concentrated using ultracentrifugation as described previously³³. Viral titre was measured using the UltraRapid Lentiviral Titer Kit (System Biosciences) and titres were described in infectious units (i.f.u.).

AAVs. Serotype 8 AAVs with Cre-recombinase-dependent flexstop cassettes that permit expression of mCherry-fused hM3Dq or mCherry-fused hM4Di under the control of the human synapsin promoter (AAV-DIO-hM3Dq-mCherry or AAV-DIO-hM4Di-mCherry) were purchased from the Vector Core at the University of North Carolina at Chapel Hill (the UNC vector core). Amount used is described in virus particles (v.p.).

Stereotaxic injection. Mice aged 2–6 months were used for injection. Viruses were injected into the brain using a Stereotaxic Alignment System (David Kopf Instruments) with an inhalation anaesthesia of 2.5% isoflurane. Virus suspensions (PRVs: 1–1.5 × 10⁶ p.f.u. (1 µl); LVF2TK: 1–1.5 × 10⁶ i.f.u. (1 µl); AAVs: 1–3 × 10⁹ v.p. (200–330 nl)) were loaded into a 1-µl syringe, and injected at 100 nl per minute. The needle was inserted to the target locations based on a stereotaxic atlas³⁰. After recovery, animals were singly housed with regular 12 h dark/light cycles, and food and water were provided *ad libitum*.

Odour exposure. Mice were exposed to a predator odour or distilled water within a modified polycarbonate vacuum-desiccator, as described previously³⁴. A single mouse was placed on a platform in a chamber of 15 cm diameter with input and output ports, exposed to charcoal-filtered air for 16 h, and then to filtered air bubbled through 12 mM 2,5-dihydro-2,4,5-trimethylthiazoline (TMT) (Contech) diluted in water, water alone, bobcat urine (Maine outdoor solutions or Murray's lures and trapping), or rabbit urine (Kishel's quality animal scents and lures). For detection of *nArc* in PRV-infected cells, CRH-Cre mice injected with PRVB177 4 days earlier were exposed to odours for 5 min. For analysis of c-Fos expression in CRH neurons, CRH-Cre mice crossed with Rosa-floxed-GFP mice were exposed to odours for 10 min and then to clean air for 50 min. All odour exposures were performed between 9:00 and 11:00 a.m.

Plasma ACTH and corticosterone assays. After mice were killed by cervical dislocation and decapitation, trunk blood was collected directly into blood collection

tubes (Becton Dickinson) containing 50 µl aprotinin (Phoenix Pharmaceuticals). Plasma was obtained by centrifugation at 1,600g for 15 min at 4°C, and stored at –80°C. Plasma ACTH concentrations were measured using the ACTH ELISA kit (MD Biosciences), according to the manufacturer's instructions, with the following modifications: (1) 100 µl of the controls or blood plasma combined with 100 µl of PBS (pH 7.4) was used in the place of 200 µl plasma; and (2) the results were assessed with the QuantaRed Enhanced Chemifluorescent HRP Substrate (Thermo Fisher). Plasma corticosterone concentrations were measured using the corticosterone ELISA kit (Abcam), according to the manufacturer's instructions, with the following modifications: (1) plasma was diluted 25 times instead of 100 times with buffer M; and (2) the results were assessed with the QuantaRed Enhanced Chemifluorescent HRP Substrate. Fluorescence was measured with a CytoFluor4000 plate reader (Applied Biosystems).

Immunofluorescence. Animals were perfused transcardially with 4% paraformaldehyde (PFA). Their brains were then soaked in 4% PFA for 4 h, in 30% sucrose for 48 h, and then frozen in OCT (Sakura) and cut into 14–20-µm coronal sections using a cryostat. Brain sections were washed twice with PBS, permeabilized with 0.5% Triton X-100 in PBS for 5 min, washed twice with PBS, blocked with TNB (Perkin Elmer) for 1 h at room temperature, and then incubated with primary antibodies diluted in TNB at 4°C overnight. Sections were then washed three times with TNT (0.1 M Tris pH 7.5, 150 mM NaCl, 0.05% Tween), incubated with the appropriate secondary antibodies and 0.5 µg ml^{–1} 4',6-diamidino-2-phenylindole (DAPI; Sigma) for 1 h at room temperature, and washed three times with TNT. Slides were coverslipped with Fluoromount-G (Southern Biotech). The following antibodies were used: (1) biotinylated mouse anti-HA (BioLegend, #901505, 1:300) followed by Alexa488-Streptavidin (Thermo Fisher, #S32354, 1:1,000) for polysynaptic PRV tracing; (2) goat anti-GFP (Rockland, #600-101-215, 1:1,000) and biotinylated anti-HA (BioLegend, 1:300) followed by Alexa488 donkey anti-goat IgG (Thermo Fisher, #A11055, 1:1,000) and Alexa555-Streptavidin (Thermo Fisher, #S32355, 1:1,000) for monosynaptic PRV tracing; (3) rabbit anti-GFP (Thermo Fisher, #A-11122, 1:500) and goat anti-c-Fos (Santa Cruz, #sc-52G, 1:300) followed by Alexa488 donkey anti-rabbit IgG (Thermo Fisher, #A21206, 1:1,000) and Alexa555 donkey anti-goat IgG (Thermo Fisher, #A21432, 1:1,000) for analysis of c-Fos expression in CRH neurons.

In situ hybridization. *In situ* hybridization was performed essentially as described previously^{11,35}, with some experiments using additional steps for dual staining. Coding region fragments of *Arc*, *Vglut1*, *Vglut2*, *Gad1*, *Gad2* and *Crh*, and the first intron sequence of *c-Fos* mRNA (for nuclear c-Fos staining) were isolated from mouse brain cDNA or mouse genomic cDNA using PCR, and cloned into the pCR4 Topo vector (Thermo Fisher). Digoxigenin (DIG)- or fluorescein (FLU)-labelled cRNA probes (riboprobes) were prepared using the DIG or FLU RNA Labelling Mix (Roche). Brains were frozen in OCT, and 16–20-µm coronal cryostat sections were hybridized to DIG- and/or FLU-labelled cRNA probes at 56°C for 13–16 h.

***nArc* mRNA staining.** After hybridization, sections were washed twice in 0.2 × SSC at 63°C for 30 min, incubated with peroxidase (POD)-conjugated anti-DIG antibodies (Roche, #11207733910, 1:2,000) at 37°C for 2 h, and then treated using the TSA-plus FLU kit (Perkin Elmer). Sections were then coverslipped with Fluoromount-G with DAPI (Southern Biotech).

Co-staining for *nArc*, *Vglut1/2* or *Gad1/2* mRNA and HA (PRVB177). After hybridization, sections were washed twice in 0.2 × SSC at 63°C for 30 min, incubated with POD-conjugated anti-DIG antibodies (Roche, #11207733910, 1:2,000) and biotinylated anti-HA antibodies (BioLegend, #901505, 1:300) at 37°C for 2 h, and then treated using the TSA-plus Cy3 kit (Perkin Elmer). Sections were then incubated with 0.5 µg ml^{–1} DAPI and Alexa488-Streptavidin (Thermo Fisher, 1:1,000) at room temperature for 1 h, and were coverslipped with Fluoromount-G.

Co-staining with *Crh* and nuclear c-Fos riboprobes. After hybridization, sections were washed twice in 0.2 × SSC at 63°C for 30 min, blocked with Streptavidin/Biotin Blocking kit (Vector Laboratories), incubated with POD-conjugated anti-FLU antibodies (Roche, #11426346910, 1:300) and alkaline phosphatase (AP)-conjugated anti-DIG antibodies (Roche, #11093274910, 1:300) at room temperature for 2 h, and then treated using the TSA-plus Biotin kit (Perkin Elmer). Sections were next incubated with 0.5 µg ml^{–1} DAPI and Alexa488-Streptavidin (Thermo Fisher, 1:1,000) at room temperature for 30 min, incubated with HNPP Fluorescent Detection Set (Roche) at room temperature for 1 h, and then coverslipped with Fluoromount-G.

Chemogenetic neuronal activation and silencing. **Activation.** AAV-DIO-hM3Dq-mCherry was injected into the pPir, AmPir or MEA of Emx1-IRES-Cre or Vglut2-IRES-Cre mice by stereotaxic injection (see earlier). At ≥2 weeks after injection, mice were intraperitoneally injected with clozapine-N-oxide (CNO; Sigma) (5.0 mg kg^{–1} body weight) or control saline. Twenty minutes later, trunk blood and brain were collected and the blood was used for plasma ACTH or corticosterone assays (see earlier). Brains were fixed with 4% PFA for 4 h, soaked in 30% sucrose

for 24 h, frozen in OCT, and cut into 20- μ m coronal sections using a cryostat. To analyse expression of hM3Dq-mCherry, brain sections were immunostained with rabbit anti-RFP (Rockland, #600-410-379, 1:500) followed by Alexa555 donkey anti-rabbit IgG (Thermo Fisher, #A31572, 1:1,000) antibodies (see earlier).

Silencing. AAV-DIO-hM4Di-mCherry was injected into the AmPir of Emx1-IRES-Cre mice by stereotaxic injection (see earlier). At ≥ 2 weeks after injection, mice were intraperitoneally injected with CNO (5.0 mg kg⁻¹ body weight) or control saline. Forty minutes later, mice were exposed to bobcat urine or TMT for 10 min (see earlier). Animal behaviour during exposure was recorded with a Canon PowerShot ELPH300HS camera placed above the chamber, and the duration that mice stayed immobile during the 10-min odour exposure were scored by using Ethovision XT11 software (Noldus Information Technology). Trunk blood and brain were then collected. Blood was used for plasma ACTH or corticosterone assays (see earlier). Brains were treated and immunostained with rabbit-anti-RFP antibodies (see earlier).

Cell counting. Images were collected using an AxioCam camera and AxioImager.Z1 microscope with apotome device (Zeiss), or TissueFAXS system (Tissuegnostics). Brain structures were identified microscopically and in digital photos using a mouse brain atlas³⁰.

To analyse the locations and number of PRV-infected cells in brain areas, every fifth section was analysed, and numbers of PRV⁺ cells were multiplied by five to acquire approximate total number of cells per animal. Brain areas were judged to contain upstream neurons if they contained ≥ 10 labelled neurons in $\geq 50\%$ of animals.

To analyse the location and number of virus-infected and *nArc*-expressing cells, three of five sections were analysed. The number of PRV⁺ cells with or without *nArc* mRNA was blindly counted. The percentages of PRV⁺ cells with only *nArc* signals among all PRV⁺ cells lacking cytoplasmic *Arc* were calculated.

To analyse the activation of CRH neurons in chemogenetic experiments, the numbers of *Crh*⁺ neurons labelled for *nc-Fos* mRNA were blindly counted.

To analyse the distribution of virus-infected cells, all sections through the regions of interest were analysed. The section images were aligned from anterior to posterior, and the numbers of virus-infected cells were counted. The size of each brain was adjusted linearly so that each posterior piriform cortex had the same length (2.8 mm).

Statistical analysis. All data are shown as the mean \pm s.e.m. Data were tested with the Shapiro–Wilk test for normality. For data with a normal distribution, the unpaired *t*-test or one-way ANOVA with post-hoc Dunnett's test was used to compare two groups or more than two groups, respectively, to analyse statistical significance. For *t*-tests, equality of variances was analysed with the *F*-test and Welch's correction was employed when variances of populations were significantly different. For data without a normal distribution, the Mann–Whitney *U*-test or the Kruskal–Wallis test with post-hoc Dunn's test was used to compare two groups or more than two groups, respectively. All tests were two-sided. Detailed information on the numbers of animals used, statistical analyses, and effect sizes is provided in Supplementary Table 1.

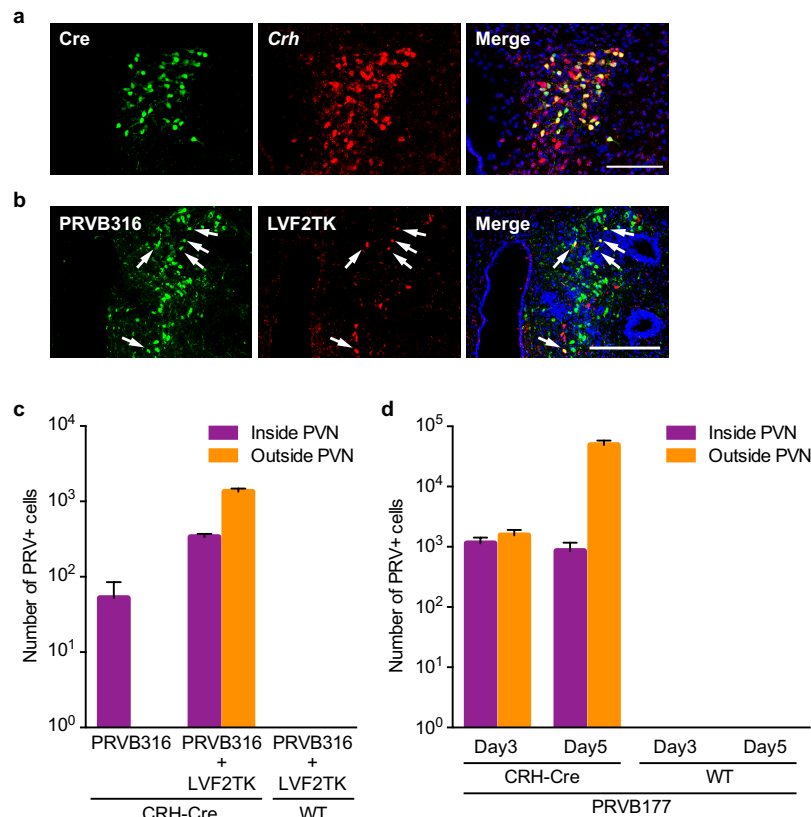
Abbreviations for brain areas. Abbreviations used for brain areas are according to Franklin and Paxinos³⁰, with a few minor modifications.

Olfactory areas. ACo, anterior cortical amygdala; AON, anterior olfactory nucleus; AmPir, amygdalo-piriform transition area; CxA, cortex-amygdala transition zone; LEnt, lateral entorhinal cortex; LOT, nucleus of lateral olfactory tract; MEAa, medial amygdala, anterior part; MEAp, medial amygdala, posterior part;

OT, olfactory tubercle; aPir, piriform cortex, anterior part; pPir, piriform cortex, posterior part; PLCo, posterolateral cortical amygdala; PMCo, posteromedial cortical amygdala; TT, tenia tecta.

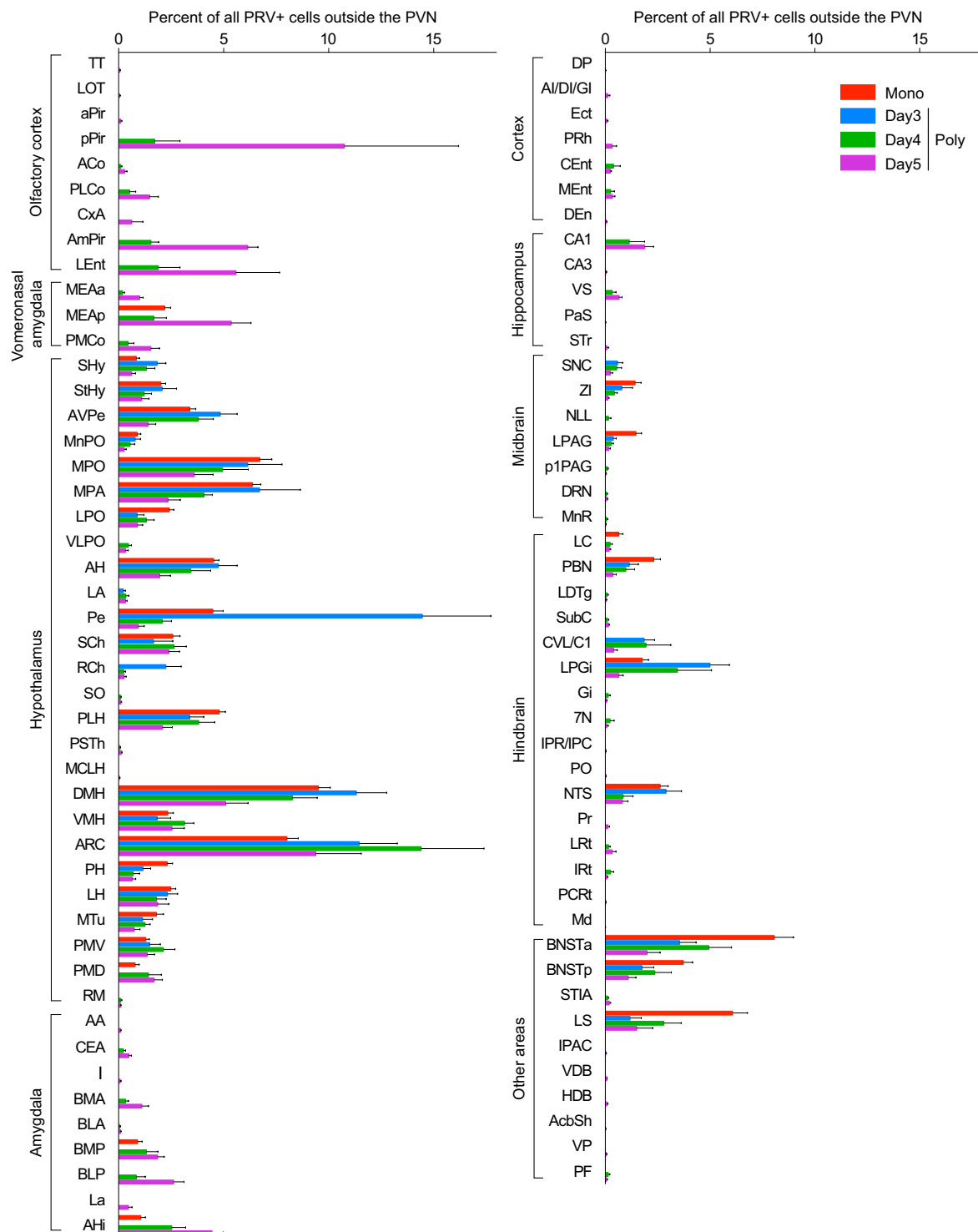
Non-olfactory areas. AA, anterior amygdaloid area; AcbSh, accumbens nucleus, shell; AH, anterior hypothalamic area; AHj, amygdalohippocampal area; AI, agranular insular cortex; ARC, arcuate hypothalamic nucleus; AVPe, anteroventral periventricular nucleus; BLA, basolateral amygdala, anterior part; BLP, basolateral amygdaloid nucleus, posterior part; BMA, basomedial amygdala, anterior part; BMP, basomedial amygdaloid nucleus, posterior part; BNSTa, bed nucleus of the stria terminalis, anterior part; BNSTp, bed nucleus of the stria terminalis, posterior part; CA1, field CA1 of the hippocampus; CA3, field CA3 of the hippocampus; CEA, central amygdala; CEnt, caudomedial entothalamic cortex; CI, caudal interstitial nucleus of the medial longitudinal; CVL, caudoventrolateral reticular nucleus; DEn, dorsal endopiriform claustrum; DI, dysgranular insular cortex; DMH, dorsomedial hypothalamic nucleus; DP, dorsal peduncular cortex; DRN, dorsal raphe nucleus; Ect, ectorhinal cortex; 7N, facial nucleus; Gi, gigantocellular reticular nucleus; HDB, nucleus of the horizontal limb of the diagonal band; I, intercalated nuclei of the amygdala; IPAC, interstitial nucleus of the posterior limb of the ant; IPC, interpeduncular nucleus, caudal subnucleus; IPR, interpeduncular nucleus, rostral subnucleus; IRt, intermediate reticular nucleus; La, lateral amygdala; LA, lateroanterior hypothalamic nucleus; LC, locus coeruleus; LDTg, laterodorsal tegmental nucleus; LH, lateral hypothalamic area; LPAG, lateral periaqueductal grey; LPGi, lateral paragigantocellular nucleus; LPO, lateral preoptic area; LRT, lateral reticular nucleus; LS, lateral septal nucleus; MCLH, magnocellular nucleus of the lateral hypothalamus; Md, medullary reticular nucleus; MEnt, medial entorhinal cortex; MnPO, median preoptic nucleus; MnR, median raphe nucleus; MPA, medial preoptic area; MPO, medial preoptic nucleus; MTu, medial tubular nucleus; NLL, nucleus of the lateral lemniscus; NTS, nucleus of the solitary tract; p1PAG, p1 periaqueductal grey; PaS, parasubiculum; PBN, parabrachial nucleus; PCRT, parvicellular reticular nucleus; Pe, periventricular nucleus of the hypothalamus; PF, parafascicular thalamic nucleus; PH, posterior hypothalamic nucleus; PLH, peduncular part of lateral hypothalamus; PMD, premammillary nucleus, dorsal part; PMV, premammillary nucleus, ventral part; PO, parolivary nucleus; Pr, prepositus nucleus; PRh, perirhinal cortex; PSTh, parasubthalamic nucleus; PVN, paraventricular nucleus of the hypothalamus; RCh, retrochiasmatic area; RM, retromammillary nucleus; SCh, suprachiasmatic nucleus; SHy, septohippocampal nucleus; SNC, substantia nigra; SO, supraoptic nucleus; StHy, striohypothalamic nucleus; STIA, bed nucleus of the stria terminalis, intraamygdaloid; STr, subiculum, transition area; SubC, subcoeruleus nucleus; VDB, vertical limb of the diagonal band; VLPO, ventrolateral preoptic nucleus; VMH, ventromedial hypothalamic nucleus; VP, ventral pallidum; VS, ventral subiculum; ZI, zona incerta.

31. Schnütgen, F. *et al.* A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nature Biotechnol.* **21**, 562–565 (2003).
32. Banfield, B. W. & Bird, G. A. Construction and analysis of alphaherpesviruses expressing green fluorescent protein. *Methods Mol. Biol.* **515**, 227–238 (2009).
33. Tiscornia, G., Singer, O. & Verma, I. M. Production and purification of lentiviral vectors. *Nature Protocols* **1**, 241–245 (2006).
34. Boehm, U., Zou, Z. & Buck, L. B. Feedback loops link odor and pheromone signaling with reproduction. *Cell* **123**, 683–695 (2005).
35. Liberles, S. D. & Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**, 645–650 (2006).



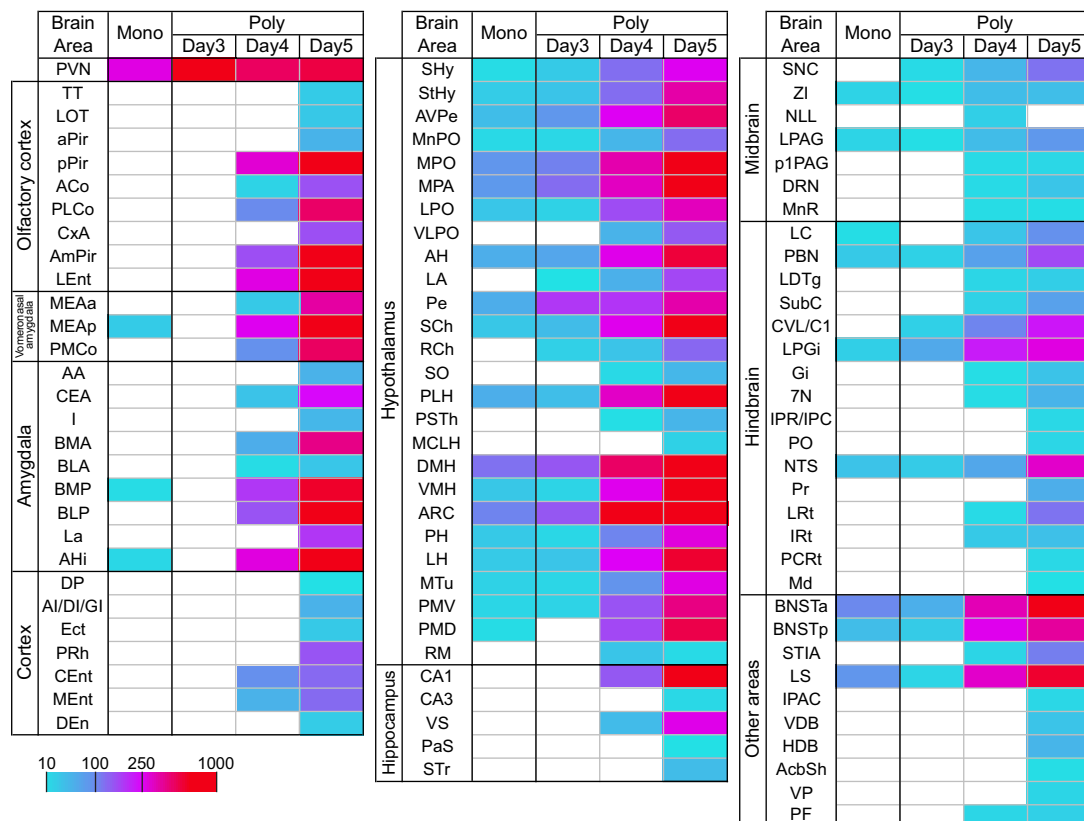
Extended Data Figure 1 | Virus infection and spread in CRH-Cre and wild-type mice. **a**, Photographs show a PVN section from CRH-Cre mice crossed with Rosa-floxstop-GFP mice that was co-stained with anti-GFP antibodies (green) (Cre) and a *Crh* riboprobe (red) (*Crh*). Co-labelled cells (yellow) can be seen in the merged image, verifying the expression of Cre in CRH neurons. Of GFP⁺ (Cre⁺) neurons in PVN sections, $93.9 \pm 0.5\%$ expressed *Crh*. $n = 4$. Scale bar, 100 μm . **b**, Photographs show a section through the PVN of a CRH-Cre mouse previously injected with PRVB316 and LVF2TK. The section was co-stained with antibodies for Cre-dependent reporters of PRVB316 (GFP (green)) and LVF2TK (HA (red)). Some co-labelled neurons are seen (arrows), verifying co-infection of some neurons by the two viruses. Scale bar, 200 μm . **c**, The number of PRVB316⁺ cells inside or outside the PVN, 4 days after PVN

injection of PRVB316 alone ($n = 2$) or PRVB316+LVF2TK ($n = 24$) in CRH-Cre mice or PRVB316+LVF2TK in wild-type (WT) mice ($n = 3$). Error bars indicate s.e.m. In CRH-Cre mice, PRVB316⁺ cells were seen both inside and outside the PVN after co-injection of PRVB316 and LVF2TK, but only inside the PVN when only PRVB316 was injected. In wild-type mice co-injected with the two viruses, PRV⁺ cells were not detected either inside or outside the PVN. **d**, The number of PRV⁺ cells inside or outside the PVN 3 or 5 days after PVN injection of PRVB177 in CRH-Cre or wild-type mice. $n = 8$ (CRH-Cre mice, day 3), $n = 6$ (CRH-Cre mice, day 5), $n = 2$ (wild-type mice, day 3), $n = 2$ (wild-type mice, day 5). Error bars indicate s.e.m. PRVB177⁺ neurons were seen both inside and outside the PVN of CRH-Cre mice on both days, but were not detected in wild-type mice at either location.



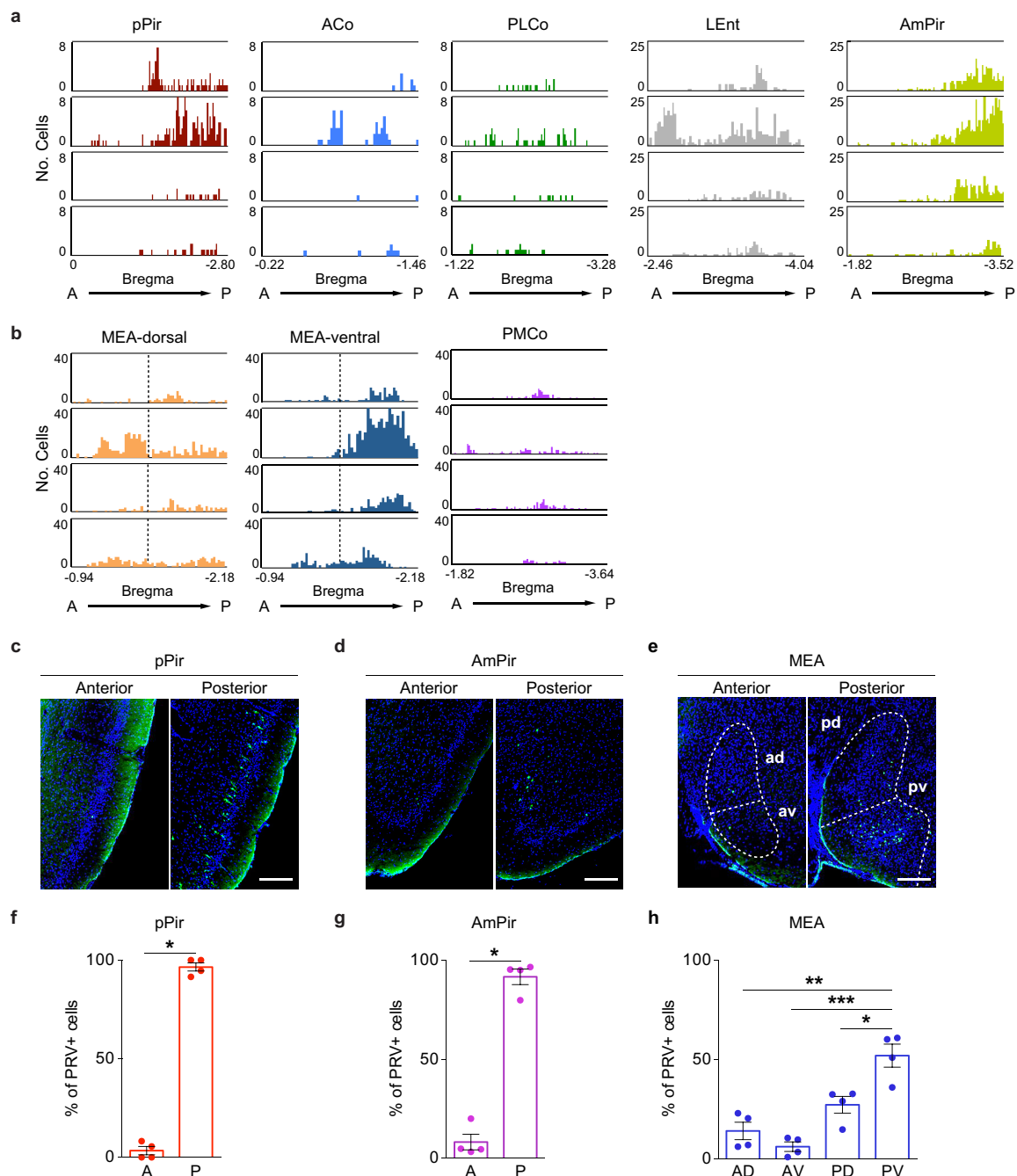
Extended Data Figure 2 | Brain areas with neurons upstream of CRH neurons. The percentage of all non-PVN PRV⁺ neurons found in individual brain areas 4 days after PVN injection of the monosynaptic PRV (PRVB316; Mono) ($n = 24$) or on day 3, 4 or 5 after PVN injection of

the polysynaptic PRV (PRVB177; Poly) ($n = 8, 6$ and 6 , for day 3, 4 and 5, respectively). Error bars indicate s.e.m. See Methods for full names of abbreviated brain areas.



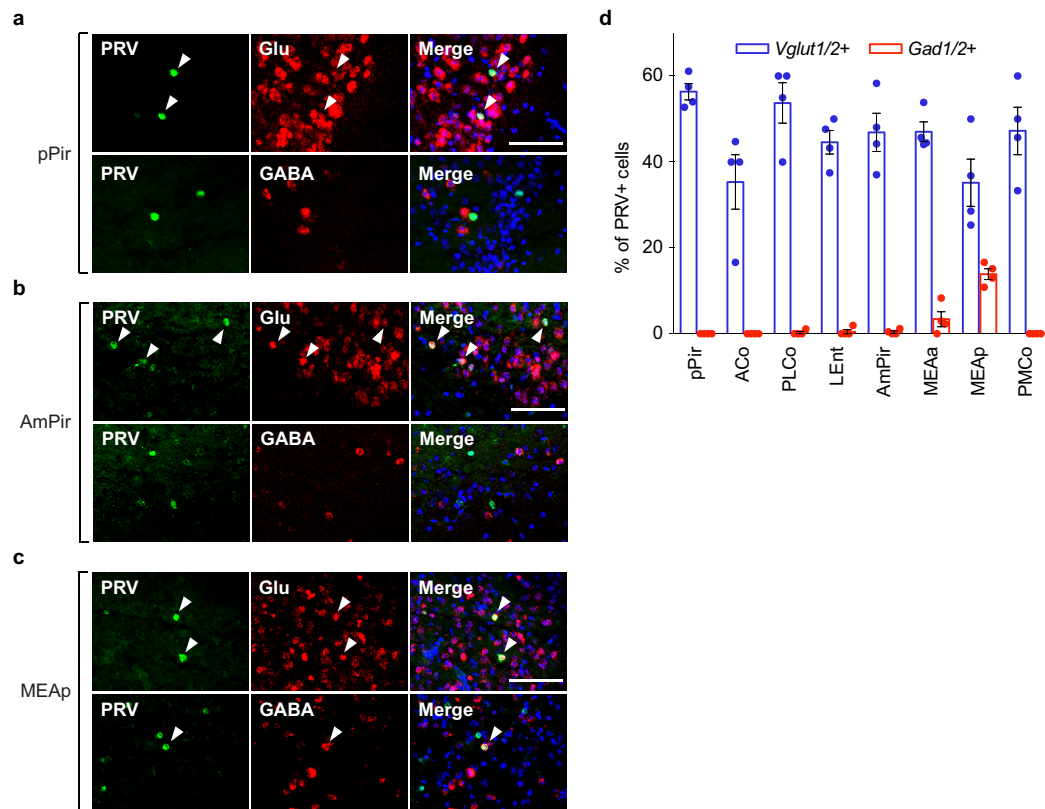
Extended Data Figure 3 | Brain areas containing neurons upstream of CRH neurons. Coloured boxes indicate the approximate number of infected neurons seen in individual brain areas (white indicates none) on day 4 after PVN injection of CRH-Cre mice with the monosynaptic

PRV (PRVB316; Mono) ($n = 24$) or on day 3, 4 or 5 after injection with the polysynaptic PRV (PRVB177; Poly) ($n = 8, 6$, and 6 , for day 3, 4 and 5, respectively). See Methods for full names of abbreviated brain areas.



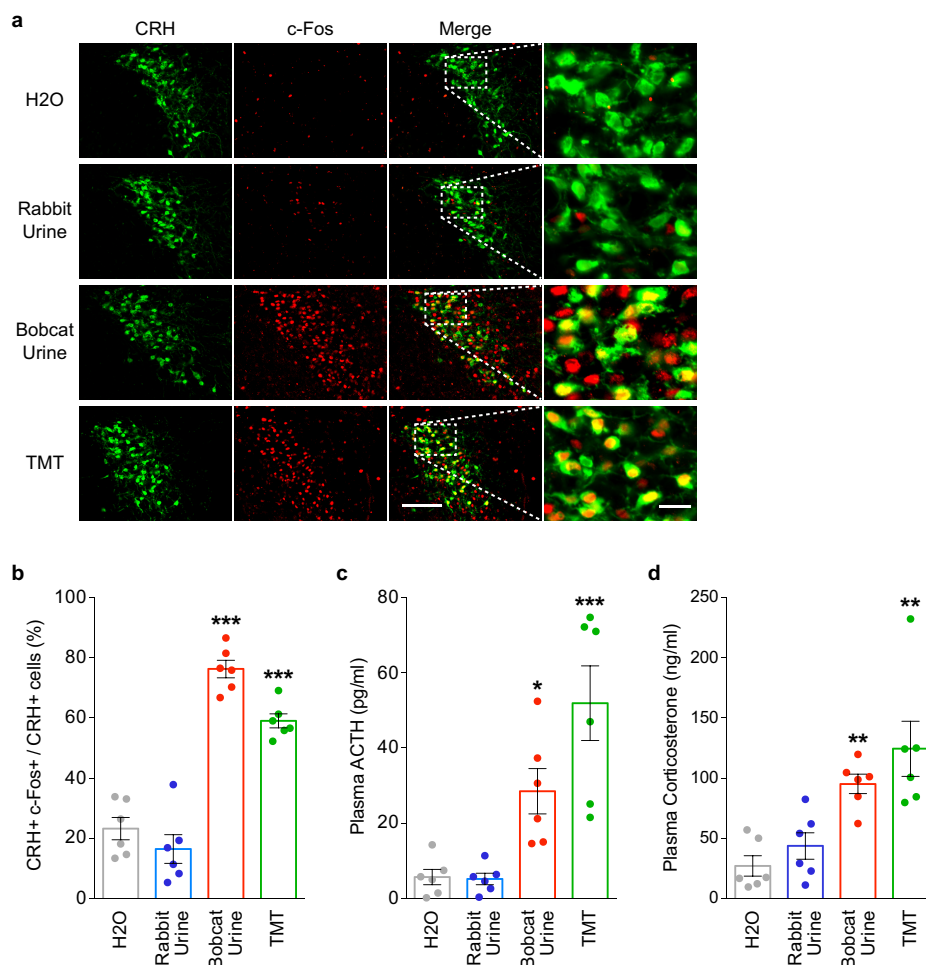
Extended Data Figure 4 | Biased distribution of neurons upstream of CRH neurons in the pPir, AmPir and MEA. **a, b**, Diagrams show the numbers of PRVB177-infected (HA⁺) neurons on day 4 after PVN injection in sequential 20-μm coronal sections along the anterior (A)–posterior (P) axis in different areas of the OC (a) and VA (b). Data are shown in separate panels for four animals for each area. Infected neurons appear to be randomly distributed along the anterior–posterior axis in some areas, but not others. **c–e**, Photographs of sections immunostained for PRVB177 (HA, green) on day 4 after PVN injection of PRVB177. Sections were counterstained with 4',6-diamidino-2-phenylindole (DAPI). More PRVB177⁺ neurons are seen in the posterior than anterior pPir (c)

and AmPir (d). In the MEA (e), more are seen in the posteroventral (pv) quadrant than other quadrants (posterodorsal (pd), anterodorsal (ad), and anteroventral (av)). Scale bars, 200 μm. **f–h**, The percentages of PRVB177-infected neurons in the anterior versus posterior halves of the pPir (f) and AmPir (g) and in different quadrants of the MEA (h). *n* = 4. Error bars indicate s.e.m. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Paired *t*-test (f, g) and paired one-way analysis of variance (ANOVA) with post-hoc Tukey's test (h). Infected neurons were concentrated in the posterior halves of the pPir and AmPir and were more abundant in the posteroventral quadrant of the MEA than in other quadrants.



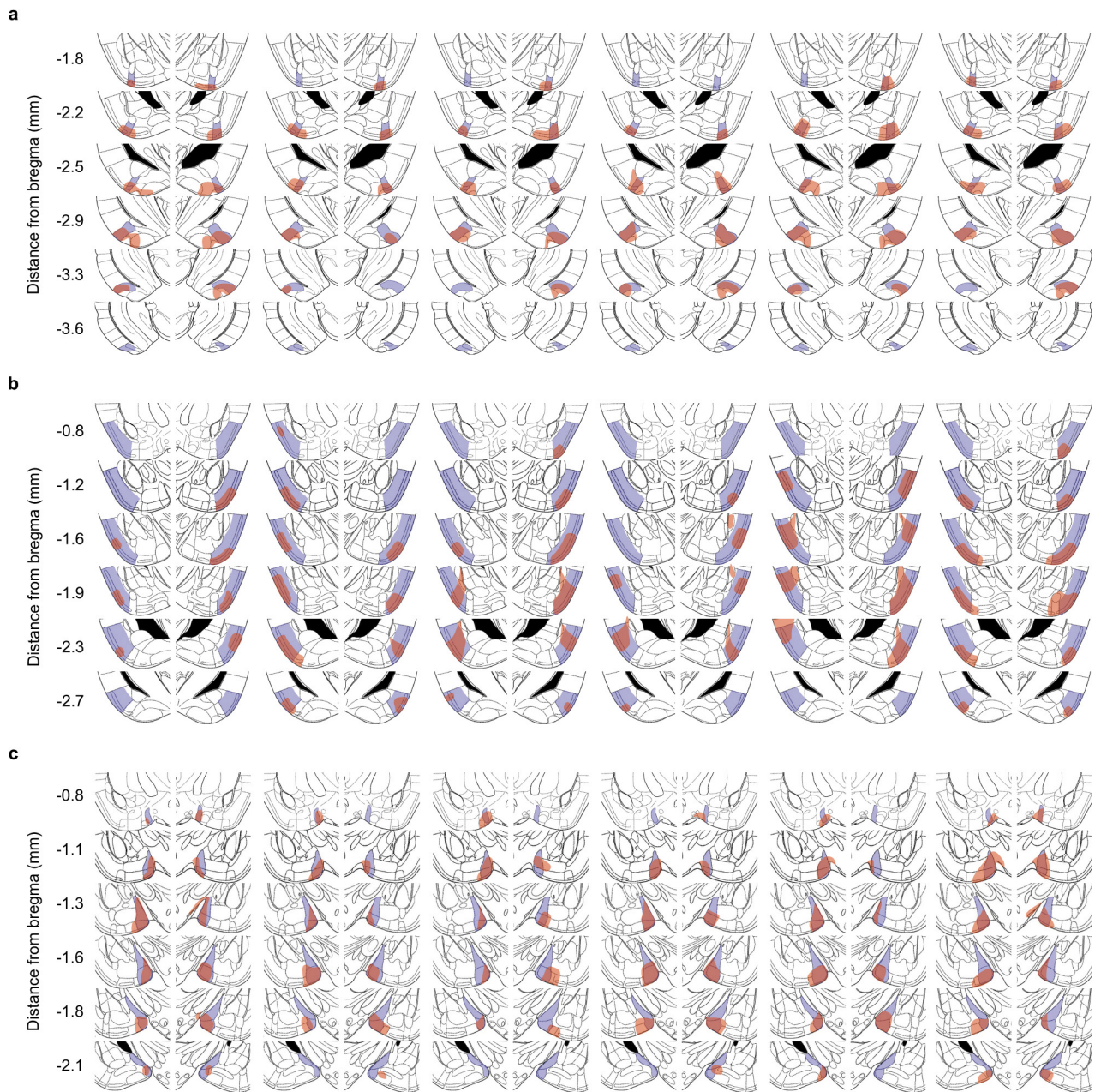
Extended Data Figure 5 | Expression of glutamatergic and GABAergic markers in PRV-infected neurons. a–c, Photographs of sections through the pPir (a), AmPir (b) or MEAp (c) 4 days after PVN injection of PRVB177. Sections were co-stained with antibodies against HA (PRVB177) (green) and riboprobes (red) for *Vglut1* and *Vglut2* (*Vglut1/2*), to identify glutamatergic neurons (Glu), or *Gad1* and *Gad2* (*Gad1/2*), to identify GABAergic neurons (GABA). Sections were counterstained with DAPI. Arrowheads indicate co-labelled neurons. Numerous PRV-infected neurons were co-labelled for *Vglut1/2* in all three areas, but

only the MEAp contained many co-labelled for *Gad1/2*. Scale bars, 100 μ m. d, The percentage of PRVB177⁺ neurons co-labelled for glutamatergic (*Vglut1/2*⁺) or GABAergic (*Gad1/2*⁺) markers in different areas of the OC and VA. PRV⁺ neurons unlabelled by either probe could express other neurotransmitters or neuromodulators or reflect PRV effects on RNA transcription or degradation. $n = 4$ mice per condition. Error bars indicate s.e.m. Most of the co-labelled PRV⁺ cells in the olfactory areas were co-labelled for the glutamatergic makers, but the MEA also contained some co-labelled for the GABAergic markers.



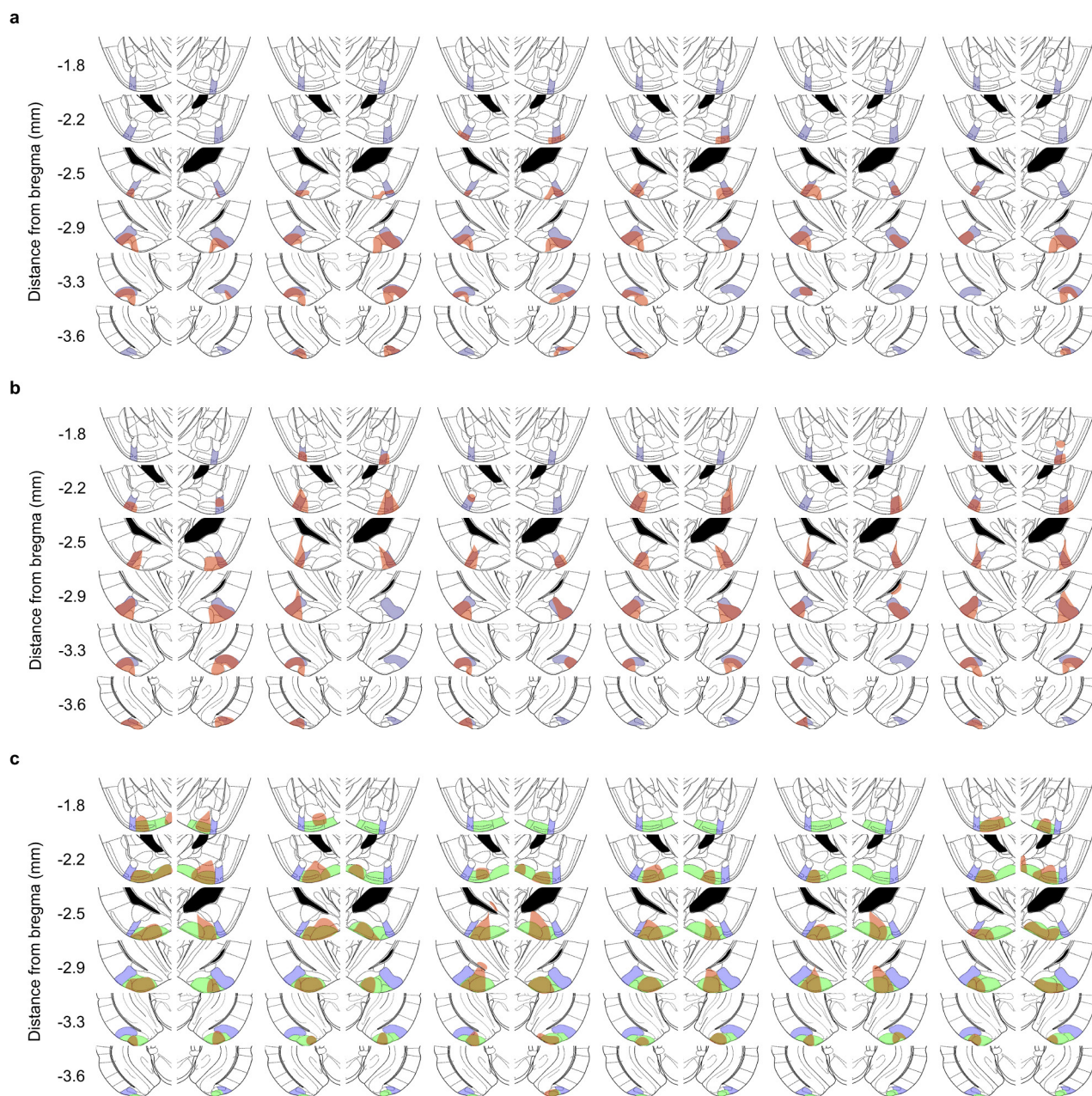
Extended Data Figure 6 | TMT and bobcat urine activate CRH neurons and induce stress hormone increases. **a**, Photographs of PVN sections from CRH-Cre mice crossed with Rosa-floxed-GFP mice. Mice were exposed to water (H₂O), rabbit urine, bobcat urine or TMT and sections were then co-stained with antibodies against GFP (green), to identify CRH neurons, and c-Fos (red), to detect activated neurons. Higher magnifications of boxed areas are shown on the right. Many CRH neurons (GFP⁺) expressed c-Fos after exposure to bobcat urine or TMT, but not rabbit urine. Scale bars, 100 μ m (left) and 20 μ m (right). **b**, The percentage of PVN CRH (GFP⁺) neurons immunostained for c-Fos after odour exposure of CRH-Cre mice crossed with Rosa-floxed-GFP mice.

Exposure to bobcat urine or TMT, but not rabbit urine, increased the percentage of CRH neurons expressing c-Fos. $n = 6$ mice per condition. Error bars indicate s.e.m. *** $P < 0.001$, one-way ANOVA with post-hoc Dunnett's test. **c**, **d**, Plasma ACTH (**c**) or corticosterone (**d**) concentrations were measured in wild-type mice after odour exposure. Exposure of animals to TMT or bobcat urine increased blood levels of ACTH and corticosterone, whereas exposure to rabbit urine did not. $n = 6$ mice per condition. Error bars indicate s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one-way ANOVA with post-hoc Dunnett's test (**c**) or Kruskal–Wallis test with post-hoc Dunn's test (**d**).



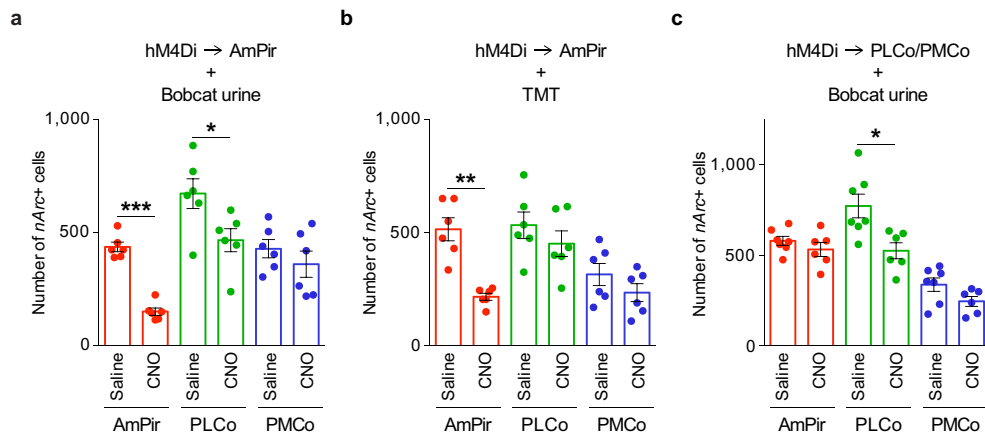
Extended Data Figure 7 | Locations of infected neurons after injection of AAV-DIO-hM3Dq-mCherry. a–c, AAV-DIO-hM3Dq-mCherry was injected into the AmPir (a), pPir (b) or MEA (c) and brain sections immunostained for mCherry to detect infected neurons. Shown are schematic sections from a mouse brain atlas³⁰ that correspond to

immunostained sections. The target area (AmPir, pPir or MEA) is indicated in purple for each section. Regions that contained the nuclei of mCherry⁺ neurons are indicated in red. Schematic sections are shown for six mice per area that were injected with CNO to test the effects of activating the area.



Extended Data Figure 8 | Locations of infected neurons after injection of AAV-DIO-hM4Di-mCherry. a–c, AAV-DIO-hM4Di-mCherry was injected into the AmPir (a, b) or PLCo/PMCo (c) and brain sections immunostained for mCherry to detect infected neurons. Shown are schematic sections from a mouse brain atlas³⁰ that correspond to immunostained sections. The location of the AmPir is indicated in purple

(a–c), and the location of the PLCo/PMCo is indicated in green (c). Regions that contained the nuclei of mCherry⁺ neurons are indicated in red in each section. Schematics are shown for six mice per area (per condition) injected with CNO and then exposed to bobcat urine (a, c) or TMT (b) to test the effects of silencing the AmPir or PLCo/PMCo.



Extended Data Figure 9 | Effect of neuronal silencing on the activation of neurons by predator odours. **a–c**, AAV-DIO-hM4Di-mCherry was injected into the AmPir (**a**, **b**) or PLCo/PMCo (**c**). Animals were later injected with CNO or saline and exposed to bobcat urine (**a**, **c**) or TMT (**b**). Brain sections were hybridized with an *Arc* riboprobe and the number of cells labelled for *nArc* was determined for the AmPir, PLCo and PMCo. Histograms show the number of cells with *nArc* in each area after injection with CNO or saline and exposure to predator odour. On the basis of

the numbers of *nArc*⁺ neurons, silencing of AmPir neurons caused a significant decrease in AmPir and PLCo neurons activated by bobcat urine (**a**) and by AmPir (but not PLCo) neurons activated by TMT (**b**). Silencing of the PLCo/PMCo (**c**) caused a significant decrease in neurons activated by bobcat urine in the PLCo, but not the AmPir or PMCo. $n=6-7$ per condition. Error bars indicate s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired t -test.

Derivation and differentiation of haploid human embryonic stem cells

Ido Sagi¹, Gloryn Chia², Tamar Golan-Lev¹, Mordecai Peretz¹, Uri Weissbein¹, Lina Sui², Mark V. Sauer³, Ofra Yanuka¹, Dieter Egli^{2,4} & Nissim Benvenisty¹

Diploidy is a fundamental genetic feature in mammals, in which haploid cells normally arise only as post-meiotic germ cells that serve to ensure a diploid genome upon fertilization. Gamete manipulation has yielded haploid embryonic stem (ES) cells from several mammalian species^{1–6}, but haploid human ES cells have yet to be reported. Here we generated and analysed a collection of human parthenogenetic ES cell lines originating from haploid oocytes, leading to the successful isolation and maintenance of human ES cell lines with a normal haploid karyotype. Haploid human ES cells exhibited typical pluripotent stem cell characteristics, such as self-renewal capacity and a pluripotency-specific molecular signature. Moreover, we demonstrated the utility of these cells as a platform for loss-of-function genetic screening. Although haploid human ES cells resembled their diploid counterparts, they also displayed distinct properties including differential regulation of X chromosome inactivation and of genes involved in oxidative phosphorylation, alongside reduction in absolute gene expression levels and cell size. Surprisingly, we found that a haploid human genome is compatible not only with the undifferentiated pluripotent state, but also with differentiated somatic fates representing all three embryonic germ layers both *in vitro* and *in vivo*, despite a persistent dosage imbalance between the autosomes and X chromosome. We expect that haploid human ES cells will provide novel means for studying human functional genomics and development.

Haploid genetics is useful for delineating genome function. Although haploid genetics has been mostly restricted to unicellular organisms, recent reports of haploid ES cells have extended it into animal species, including mammals^{1–7}. Deriving haploid human ES cells by similar approaches has probably been hindered by the limited availability of human oocytes⁸. Artificial activation of unfertilized metaphase II (MII) human oocytes results in efficient development to the blastocyst stage and subsequent derivation of parthenogenetic ES (pES) cells^{9–11}. In mouse parthenogenetic embryos haploidy usually persists at the blastocyst stage^{12,13}, but diploid cells progressively dominate over increasing cell cycles due to spontaneous and irreversible diploidization^{13–15} (Fig. 1a). By estimation, even if diploidization occurs in 1 out of 10 cell cycles, 1% of ES cells may remain haploid at early passages (Extended Data Fig. 1a).

To explore the feasibility of deriving haploid human ES cells, we generated and analysed a collection of 14 early-passage human pES cell lines for the persistence of haploid cells. We initially used chromosome counting by metaphase spreading as a method for unambiguous and quantitative discovery of rare haploid nuclei. Among ten individual pES cell lines, a low proportion of haploid metaphases was found in one cell line, pES10 (1.3%, Extended Data Table 1). In four additional lines, we also used viable fluorescence-activated cell sorting (FACS) with Hoechst 33342 staining, aiming to isolate cells with a DNA content

corresponding to less than two chromosomal copies (2c), leading to the successful enrichment of haploid cells from a second cell line, pES12 (Extended Data Table 2).

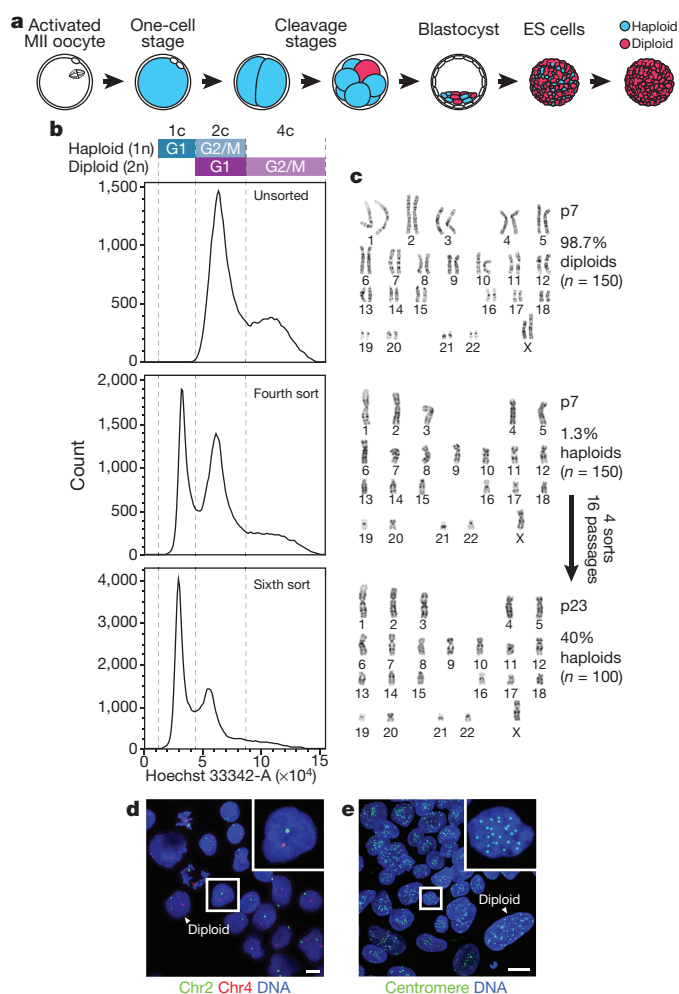


Figure 1 | Derivation of haploid human ES cells. **a**, Schematic of putative haploidy in pES cells. **b**, DNA content profiles of haploid pES10, established by repeated enrichment of 1c cells. Top to bottom, unsorted diploid cells, partially purified (fourth sort) and mostly purified (sixth sort) haploid cells. **c**, pES10 karyotypes before and after 1c-cell enrichment. **p**, passage. **d**, **e**, DNA FISH (**d**) and centromere staining (**e**) in haploid-enriched pES10 cells. Magnifications show haploid nuclei with a single hybridization signal (**d**) and 23 centromere foci (**e**), respectively. Scale bars, 10 μ m.

¹The Azrieli Center for Stem Cells and Genetic Research, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel. ²Department of Pediatrics, Columbia University, New York, New York 10032, USA. ³Center for Women's Reproductive Care, College of Physicians and Surgeons, Columbia University, New York, New York 10019, USA. ⁴The New York Stem Cell Foundation Research Institute, New York, New York 10032, USA.

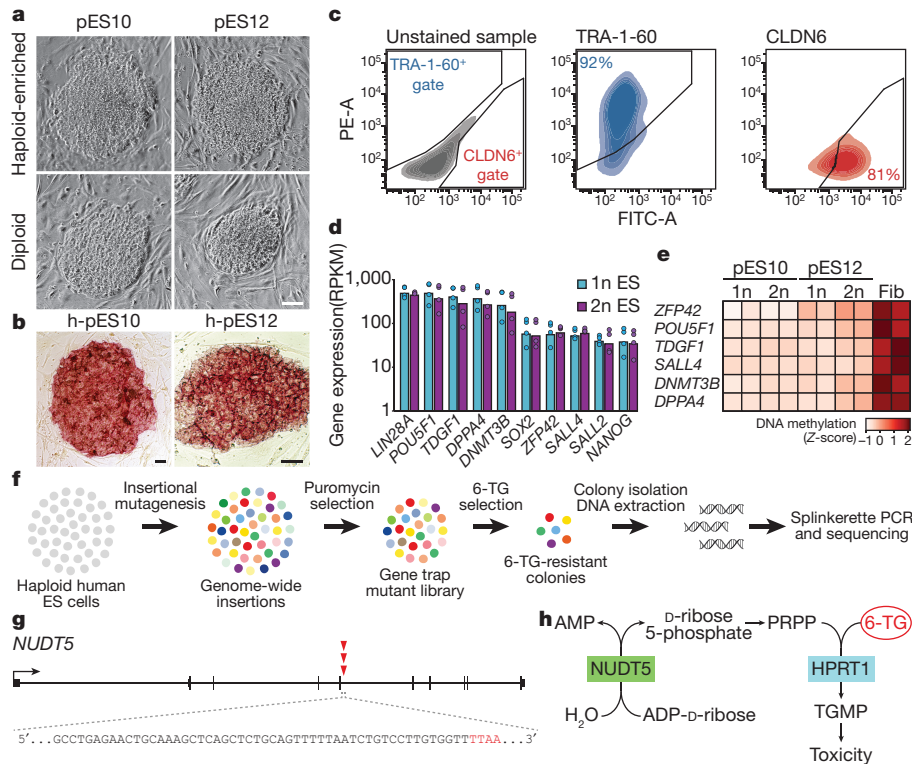


Figure 2 | Haploid human ES cells display pluripotent stem cell characteristics and enable loss-of-function screening. **a, b**, Colony morphology and alkaline phosphatase staining. Scale bars, 50 μ m. **c**, Flow cytometry analysis of gated h-pES10 1c cells by co-staining DNA and cell surface markers TRA-1-60 and CLDN6. **d**, Expression levels of pluripotency genes in G1-sorted haploid and diploid ES cells ($n = 4$ each,

two biological replicates per line, logarithmic scale). Mean indicated by bars. **e**, DNA methylation levels at pluripotency genes in biological duplicates of G1-sorted ES cells and control fibroblasts (Fib). **f**, Schematic of gene trapping and screening for 6-TG-resistance genes. **g**, *NUDT5* insertions (red arrows) detected in three 6-TG-resistant colonies (indicated by TTA). **h**, Metabolic pathway leading to 6-TG toxicity.

Two individual haploid-enriched ES cell lines were established from pES10 and pES12 (hereafter referred to as h-pES10 and h-pES12) within five to six rounds of 1c-cell enrichment and expansion (Fig. 1b and Extended Data Fig. 1b). These cell lines were cultured in standard conditions for over 30 passages while including cells with a normal haploid karyotype (Fig. 1c and Extended Data Fig. 1c). However, since diploidization occurred at a rate of 3–9% cells per day (Extended Data Fig. 1a; see Methods), sorting at every three to four passages was required to maintain haploid cells. Visualization of ploidy was further enabled by DNA fluorescence *in situ* hybridization (FISH) (Fig. 1d and Extended Data Fig. 1d) and quantification of centromere protein foci (Fig. 1e and Extended Data Fig. 1e; see Supplementary Notes and Extended Data Fig. 2). Besides having an intact karyotype, haploid ES cells did not harbour significant copy number variations (CNVs) relative to their unsorted diploid counterparts (Extended Data Fig. 1f), nor common duplications that would result in pseudo-diploidy, indicating that genome integrity was preserved throughout haploid-cell isolation and maintenance.

Both h-pES10 and h-pES12 exhibited classical human pluripotent stem cell features, including typical colony morphology and alkaline phosphatase activity (Fig. 2a, b). Single haploid ES cells expressed hallmark pluripotency markers as confirmed by centromere foci quantification in essentially pure (>95%) haploid cultures (Extended Data Fig. 3). Selective flow cytometry enabled the validation of the expression of two human ES-cell-specific cell surface markers (TRA-1-60 and CLDN6 (ref. 16)) in single haploid cells (Fig. 2c). Moreover, sorted haploid and diploid ES cells showed highly similar transcriptional and epigenetic signatures of pluripotency genes (Fig. 2d, e). In accordance with a parthenogenetic origin, these cells also featured distinct molecular profiles of maternal imprinting (Extended Data Fig. 4a–d).

Haploid cells are valuable for genetic screening because phenotypically selectable mutants can be identified upon disruption of single-copy alleles. To demonstrate the applicability of this principle in haploid human ES cells, we generated a genome-wide mutant library using a gene trap transposon system (Fig. 2f and Extended Data Fig. 4e; see Methods), and screened for resistance to the purine analogue 6-thioguanine (6-TG). Out of six isolated and analysed 6-TG-resistant colonies, three harboured an identical gene trap insertion at the *NUDT5* autosomal gene (Fig. 2g). The disruption of this gene was recently confirmed to confer 6-TG resistance in human cells¹⁷ (Fig. 2h). Detection of a loss-of-function phenotype due to an autosomal mutation thereby validates that genetic screening is feasible in haploid human ES cells.

The ability of human ES cells to exist as both haploids and diploids led us to investigate whether these two ploidy states differ in certain aspects of gene regulation and cell biology. To analyse haploid and diploid ES cells in the same cell cycle phase, we used FACS to isolate G1-phase haploid cells (1c) and compared them with isogenic G1-phase diploid cells (2c) from unsorted diploid cultures (Fig. 3a and Extended Data Fig. 5a; see Methods). We first aimed to uncover putative ploidy-associated differences by comparing the transcriptomes of haploid and diploid ES cells using RNA sequencing (RNA-seq), considering that observed changes in expression levels would be relative to the total gene expression of each ploidy state, rather than representing absolute differences. On the genome-scale, undifferentiated haploid and diploid ES cells clustered closely and separately from differentiated embryoid bodies (EBs) (Fig. 3b). Nonetheless, we identified 275 relatively upregulated and 290 relatively down-regulated genes in haploids compared with diploids (greater than twofold change, false discovery rate (FDR) <0.05; Extended Data Fig. 5b).

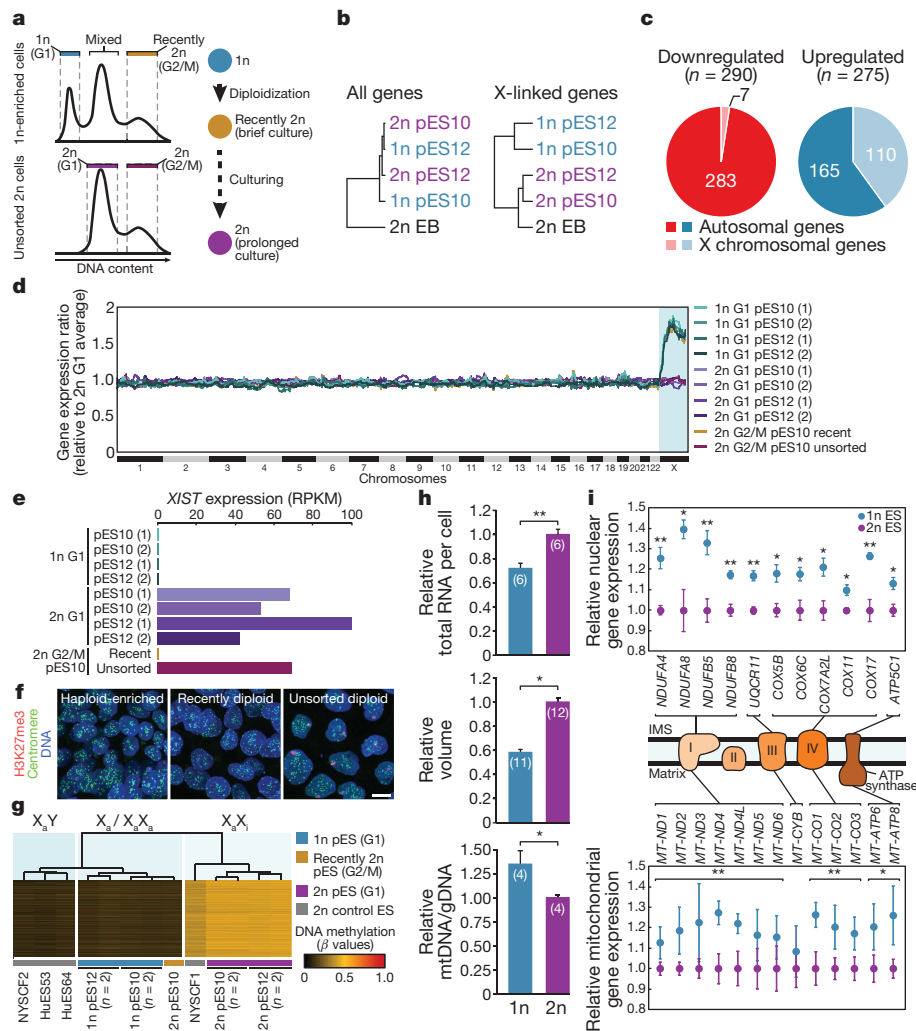


Figure 3 | Molecular and cellular comparisons of haploid and diploid ES cells. **a**, Experimental scheme. **b**, RNA-seq-based clustering analyses of isogenic G1-sorted haploid and diploid cells (two biological replicates per line) and diploid pES12-derived EBs. **c**, Pie chart of relatively downregulated and upregulated genes in haploid versus diploid ES cells. **d–g**, Differential X chromosome inactivation in haploid and diploid ES cells. **d**, Genome-wide expression moving median plot. **e**, *XIST* expression levels. **f**, H3K27me3 staining. Scale bar, 10 μ m. **g**, X chromosome DNA

methylation levels. **h**, Relative total RNA, cell volume and ratio of mitochondrial to genomic DNA (mtDNA/gDNA) between G1-sorted haploid and diploid ES cells. Number of replicates indicated in parenthesis. Error bars represent s.d. **i**, Mean expression levels \pm s.e.m. of nuclear and mitochondrial oxidative phosphorylation genes in haploid and diploid ES cells (replicates as in Fig. 2d), and their schematic organization in this pathway. * $P < 0.05$; ** $P < 0.01$ (two-tailed unpaired Student's *t* test). Source data for **h** and **i** are available online.

Notably, X chromosomal genes were significantly overrepresented among the relatively upregulated gene set (40%, $P < 0.001$, χ^2 goodness of fit test) (Fig. 3c), and the expression levels of X chromosomal genes alone clearly distinguished between haploid and diploid ES cells (Fig. 3b). These data correlate with an expected differential status of X chromosome inactivation in haploid and diploid human ES cells: whereas the single-copy X chromosome in haploids is transcriptionally active (X_a), one of the two X chromosomes in diploids is often inactivated (X_aX_i)¹⁸. Indeed, haploid human ES cells exhibited a relative increase in X chromosomal gene expression and lacked expression of the *XIST* transcript which drives X chromosome inactivation (Fig. 3d, e and Extended Data Fig. 5b–d), as in diploid X_aX_a human ES cells¹⁹. X chromosome inactivation is regulated by repressive histone modifications and DNA methylation. H3K27me3 foci were consistently observed in unsorted diploid ES cells, but not in the haploid-enriched counterparts (Fig. 3f). Moreover, methylome analysis showed that the X chromosome DNA methylation signature of haploid ES cells resembles that of diploid male ES cells (X_aY), whose single-copy X chromosome is largely hypomethylated, as opposed to the composite pattern of a hypomethylated X_a and a hypermethylated X_i in diploid female cells

(Fig. 3g). Interestingly, recently diploidized ES cells (see Methods) remained X_aX_a soon after diploidization by all assays mentioned earlier (Fig. 3a, d–g).

Normalization to total gene expression²⁰ resulted in seemingly similar expression levels of autosomal genes, but higher levels of X-linked genes in haploid cells (Fig. 3d and Extended Data Fig. 5c). However, assuming that the absolute expression of X-linked genes in haploid X_a and diploid X_aX_i cells are equivalent, these data suggest a genome-wide autosomal gene expression level reduction in haploids (Extended Data Fig. 5e, f). In support of this, total RNA amounts isolated from haploid ES cells were significantly lower than those from equal numbers of diploid cells (Fig. 3h). A decrease in total gene expression implied that physical dimensions may also be altered. Indeed, the average haploid:diploid diameter ratio of G1-sorted ES cells was ~ 0.8 , corresponding to ~ 0.7 surface area ratio and ~ 0.6 volume ratio (Fig. 3h and Extended Data Fig. 5g).

We subsequently focused on consistent differential regulation within autosomes (see Methods). By transcriptome and methylome analyses, genes relatively downregulated in haploid ES cells were significantly enriched for genes encoding signal-peptide-bearing

(Fig. 3g). Interestingly, recently diploidized ES cells (see Methods) remained X_aX_a soon after diploidization by all assays mentioned earlier (Fig. 3a, d–g).

Normalization to total gene expression²⁰ resulted in seemingly similar expression levels of autosomal genes, but higher levels of X-linked genes in haploid cells (Fig. 3d and Extended Data Fig. 5c). However, assuming that the absolute expression of X-linked genes in haploid X_a and diploid X_aX_i cells are equivalent, these data suggest a genome-wide autosomal gene expression level reduction in haploids (Extended Data Fig. 5e, f). In support of this, total RNA amounts isolated from haploid ES cells were significantly lower than those from equal numbers of diploid cells (Fig. 3h). A decrease in total gene expression implied that physical dimensions may also be altered. Indeed, the average haploid:diploid diameter ratio of G1-sorted ES cells was ~ 0.8 , corresponding to ~ 0.7 surface area ratio and ~ 0.6 volume ratio (Fig. 3h and Extended Data Fig. 5g).

We subsequently focused on consistent differential regulation within autosomes (see Methods). By transcriptome and methylome analyses, genes relatively downregulated in haploid ES cells were significantly enriched for genes encoding signal-peptide-bearing

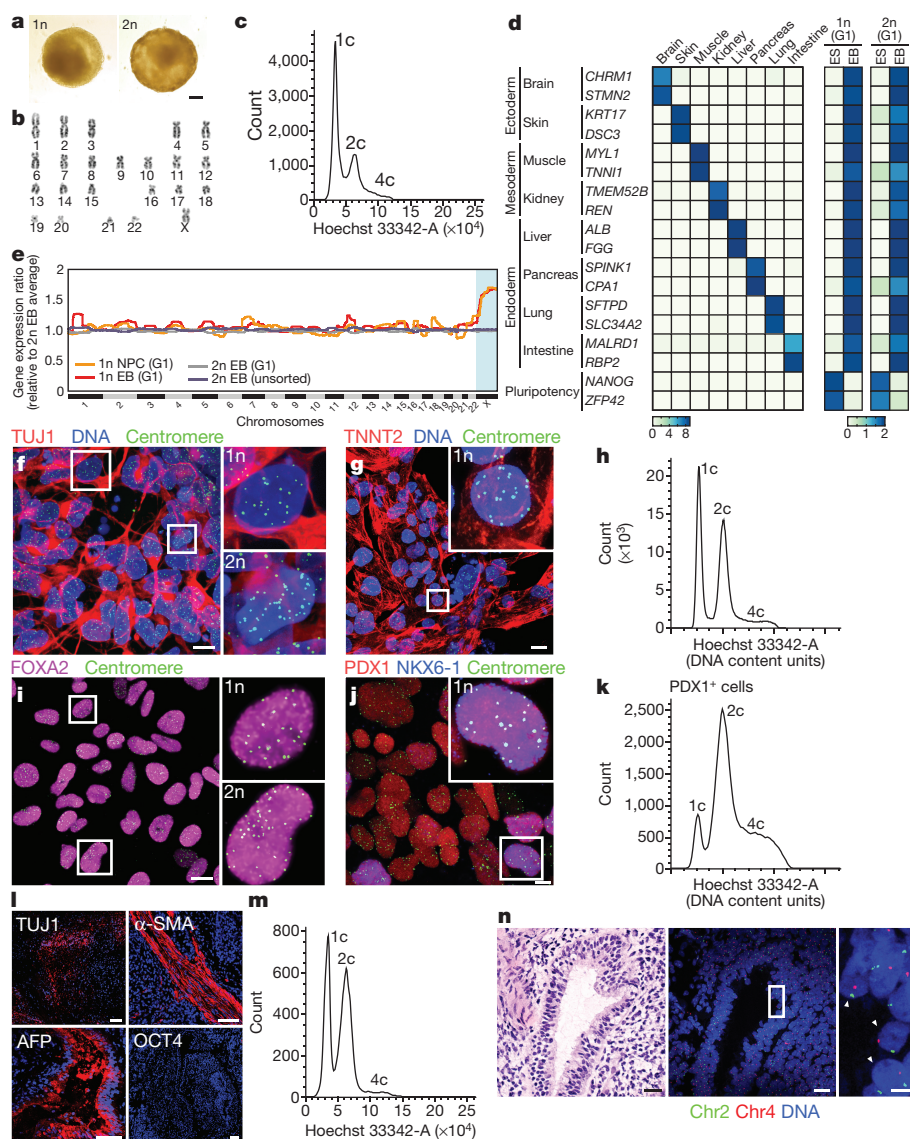


Figure 4 | Differentiation of haploid human cells. **a**, EBs from haploid-enriched and diploid pES12 cells. Scale bar, 100 μ m. **b**, Haploid EB cell karyotype. **c**, DNA content profile of h-pES10 EB cells. **d**, Expression of tissue- and pluripotency-specific genes in G1-sorted haploid and diploid ES and EB pES10 cells. **e**, Differential X chromosome inactivation in haploid and diploid EBs and NPCs by genome-wide expression moving median plot. **f**, **g**, **i**, **j**, Centromere and differentiation marker co-staining in h-pES12-derived neurons (**f**), cardiomyocytes (**g**), definitive endoderm cells (**i**) and pancreatic cells (**j**). Scale bars, 10 μ m. **h**, **k**, DNA content profiles of h-pES12 cells differentiated into cardiomyocytes (**h**) and PDX1-positive pancreatic cells (**k**). **l**, TUJ1 (ectoderm), α -SMA (mesoderm), AFP (endoderm) and OCT4 (pluripotency) staining in an h-pES12-derived teratoma. Scale bars, 50 μ m. **m**, DNA content profile of an h-pES10-derived teratoma. **n**, Serial h-pES12-derived teratoma sections analysed histologically (left; scale bar, 20 μ m) and by DNA FISH (middle; scale bar, 20 μ m). Haploid nuclei are shown in magnification (right; scale bar, 5 μ m).

proteins (Extended Data Fig. 5h). Remarkably, we also detected subtle yet significant relative upregulation of 11 genes involved in oxidative phosphorylation in haploid cells, including representatives encoding subunits of four out of the five complexes comprising this pathway (Fig. 3i and Extended Data Fig. 5i). Furthermore, all 13 mitochondrially encoded oxidative phosphorylation genes were consistently upregulated as well (Fig. 3i), indicating coordinated regulation between nuclear and mitochondrial genes. This coincided with a 32% increase in the mitochondrial DNA (mtDNA) to nuclear DNA ratio between haploids and diploids (Fig. 3h), suggesting that mitochondrial abundance relative to nuclear DNA content is higher in haploid cells.

We next sought to assess the differentiation potential of haploid human ES cells. The 21-day-old EBs generated by spontaneous differentiation of haploid-enriched and diploid ES cells could not be distinguished by their appearance (Fig. 4a), and the morphology of dissociated haploid-cell-derived EB cells was consistent with differentiation (Extended Data Fig. 6a). Notably, metaphase spreading revealed a haploid karyotype (Fig. 4b; 4/4 metaphases), and a largely haploid ($\sim 70\%$) DNA profile was confirmed by flow cytometry in both h-pES10-derived and h-pES12-derived EB cells (Fig. 4c and Extended Data Fig. 6b). We then compared the gene expression profiles of G1-sorted haploid ES cells and EB cells, focusing on 18 lineage-specific genes across nine cell types (Fig. 4d). Whereas expression levels were

negligible in undifferentiated ES cells, all tissue-specific genes were expressed in haploid and diploid EB cells (Fig. 4d and Extended Data Fig. 6c). Haploid and diploid EB cells showed insignificant expression of pluripotency-specific genes, consistent with efficient acquisition of differentiated cell fates.

To extend this analysis to more specific cell types, we subjected haploid ES cells to directed differentiation assays. Haploid ES cells differentiated towards a neural fate for ten days remained haploid, while giving rise to NCAM1-positive neural progenitor cells (NPCs, with $\sim 90\%$ efficiency) (Extended Data Fig. 7a, b). Sorted haploid NPCs expressed multiple neural-lineage-specific genes but not pluripotency-specific genes (Extended Data Fig. 7c, d). X chromosome inactivation in differentiated diploid female cells results in dosage compensation of 1:2 between the X chromosome and autosomes. As haploid cells do not inactivate their single-copy X chromosome, an X:autosomes dosage imbalance of 1:1 should persist into the differentiated state. Indeed, both haploid NPCs and EB cells showed an X_a signature contrary to the X_aX_i signature of diploid EB cells as indicated by whole-genome expression analysis and *XIST* levels (Fig. 4e and Extended Data Fig. 7e).

Neuronal differentiation was not restricted to the progenitor stage as the cells also differentiated with high efficiency ($>90\%$) into mature TUJ1 (also known as β -tubulin III)-positive neurons by 20 days, with persistence of haploid cells as shown by co-staining

with centromeres (Fig. 4f; 47% haploids, $n = 104$) and FISH analysis (Extended Data Fig. 7f, g; 46% haploids, $n = 200$). Similarly, haploid cells differentiated into TNNT2-expressing cardiomyocytes (Fig. 4g; 32% haploids, $n = 97$) during an eleven-day protocol resulting in spontaneously beating clusters (Supplementary Video 1), and 39% ($n = 31$) of haploid cells sorted from the whole culture (25% 1c-cells) were confirmed as TNNT2-positive (Fig. 4h and Extended Data Fig. 7h). Next, we differentiated haploid-enriched cultures (~70% haploids) to the pancreatic lineage. Analysing two stages of differentiation by centromere foci analysis, we observed robust differentiation (>90%) of both haploids and diploids into FOXA2-positive definitive endoderm cells (Fig. 4i; 56% haploids, $n = 112$), and differentiation into PDX1-positive pancreatic cells (Fig. 4j; 13% haploid, $n = 103$), some of which were also positive for NKX6-1. In addition, the persistence of haploid PDX1-positive cells was confirmed by flow cytometry (Fig. 4k; 10% PDX1-positive 1c cells; and Extended Data Fig. 7i, j).

Finally, both haploid-enriched human ES cell lines gave rise to teratomas comprising cell types of ectodermal, mesodermal and endodermal origins (Fig. 4l and Extended Data Fig. 8a, b), meeting the most stringent criterion for human pluripotency *in vivo*. Importantly, no residual undifferentiated OCT4-positive cells were detected (Fig. 4l and Extended Data Fig. 8b). Upon dissection, DNA content analysis revealed that a considerable population of h-pES10-derived teratoma cells remained haploid (Fig. 4m). Combined analysis of serial sections from an independent, h-pES12-derived teratoma, by histology and FISH confirmed the existence of *in vivo* differentiated haploid human cells capable of contributing to an organized tissue structure while responding to developmental signals (Fig. 4n). Haploid cells were identified in all analysed teratomas ($n = 4$), albeit with variable proportions, which may be influenced by the initial amount of haploid cells and/or the duration of teratoma formation.

Haploid mammalian cells have proven invaluable for loss-of-function screens⁷. Using a genome-wide library of gene-trapped haploid human ES cells, we demonstrated their potential for biomedically relevant functional genomics by forward genetic screening (Supplementary Discussion). Whereas previous studies on non-human haploid ES cells mostly emphasized the similarity between haploids and diploids, here we also pointed to several transcriptional, epigenetic and physical properties that set them apart (Supplementary Discussion). Interestingly, we did not observe global transcriptional compensation in haploid cells, indicating that it is not required for cellular viability as long as an autosomal balance is preserved. In contrast, autosomal imbalance appears intolerable based on the strict absence of human autosomal monosomies in *in vitro* fertilization ES cells²¹. Remarkably, we found that a haploid human karyotype is not a barrier for ES cell differentiation. As observed in the mouse², haploid human ES cells gave rise to NPCs while remaining haploid. However, while mouse studies showed that haploid cells are lost upon further differentiation^{2,14}, we observed specification of human haploid cells into somatic cell fates of all three embryonic germ layers, despite persistent dosage imbalance between the X chromosome and autosomes (Supplementary Discussion).

Throughout evolution, mammalian genomes have been solidified by diploidy-dependent adaptations such as parental imprinting, which restrict the development of haploid uniparental embryos. Nonetheless, haploid cells are capable of directing development in certain animal species²². The surprising differentiation potential of haploid human genomes suggests that diploidy-dependent adaptations, rather than haploidy, pose the predominant barriers for uniparental development in humans. The discovery of haploid human ES cells should thus provide novel means to delineate basic aspects of human genetics and development.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 July 2015; accepted 8 February 2016.

Published online 16 March 2016.

1. Leeb, M. & Wutz, A. Derivation of haploid embryonic stem cells from mouse embryos. *Nature* **479**, 131–134 (2011).
2. Elling, U. *et al.* Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9**, 563–574 (2011).
3. Yang, H. *et al.* Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. *Cell* **149**, 605–617 (2012).
4. Li, W. *et al.* Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* **490**, 407–411 (2012).
5. Li, W. *et al.* Genetic modification and screening in rat using haploid embryonic stem cells. *Cell Stem Cell* **14**, 404–414 (2014).
6. Yang, H. *et al.* Generation of haploid embryonic stem cells from *Macaca fascicularis* monkey parthenotes. *Cell Res.* **23**, 1187–1200 (2013).
7. Wutz, A. Haploid mouse embryonic stem cells: rapid genetic screening and germline transmission. *Annu. Rev. Cell Dev. Biol.* **30**, 705–722 (2014).
8. Egli, D. *et al.* Impracticability of egg donor recruitment in the absence of compensation. *Cell Stem Cell* **9**, 293–294 (2011).
9. Revazova, E. S. *et al.* Patient-specific stem cell lines derived from human parthenogenetic blastocysts. *Cloning Stem Cells* **9**, 432–449 (2007).
10. Kim, K. *et al.* Recombination signatures distinguish embryonic stem cells derived by parthenogenesis and somatic cell nuclear transfer. *Cell Stem Cell* **1**, 346–352 (2007).
11. Paull, D. *et al.* Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* **493**, 632–637 (2013).
12. Tarkowski, A. K., Witkowska, A. & Nowicka, J. Experimental parthenogenesis in the mouse. *Nature* **226**, 162–165 (1970).
13. Kaufman, M. H., Robertson, E. J., Handyside, A. H. & Evans, M. J. Establishment of pluripotential cell lines from haploid mouse embryos. *J. Embryol. Exp. Morphol.* **73**, 249–261 (1983).
14. Leeb, M. *et al.* Germline potential of parthenogenetic haploid mouse embryonic stem cells. *Development* **139**, 3301–3305 (2012).
15. Takahashi, S. *et al.* Induction of the G2/M transition stabilizes haploid embryonic stem cells. *Development* **141**, 3842–3847 (2014).
16. Ben-David, U., Nudel, N. & Benvenisty, N. Immunologic and chemical targeting of the tight-junction protein Claudin-6 eliminates tumorigenic human pluripotent stem cells. *Nat. Commun.* **4**, 1992 (2013).
17. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnol.* **34**, 184–191 (2016).
18. Silva, S. S., Rowntree, R. K., Mekhoubad, S. & Lee, J. T. X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 4820–4825 (2008).
19. Bruck, T., Yanuka, O. & Benvenisty, N. Human pluripotent stem cells with distinct X inactivation status show molecular and cellular differences controlled by the X-linked ELK-1 gene. *Cell Rep.* **4**, 262–270 (2013).
20. Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).
21. Biancotti, J. C. *et al.* The *in vitro* survival of human monosomies and trisomies as embryonic stem cells. *Stem Cell Res.* **9**, 218–224 (2012).
22. Otto, S. P. & Jarne, P. Evolution. Haploids-hapless or happening? *Science* **292**, 2441–2443 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank all members of the Benvenisty and Egli laboratories for input and support. We thank Y. Avior and W. Breuer for their assistance with experimental procedures. I.S. is supported by the Adams Fellowships Program for Doctoral Students, G.C. is supported by the A*STAR International Fellowship, U.W. is a Clore Fellow, D.E. is a NYSCF-Robertson Investigator, and N.B. is the Herbert Cohn Chair in Cancer Research. This work was partially supported by The Rosetrees Trust and by The Azrieli Foundation (to N.B.), by the Russell Berrie Foundation Program in Cellular Therapies of Diabetes, by the New York State Stem Cell Science (NYSTEM) IIRP Award number CQ26184, and by the New York Stem Cell Foundation (to D.E.).

Author Contributions I.S., D.E. and N.B. designed the study and wrote the manuscript with input from all authors. I.S. isolated and characterized haploid human ES cell lines, performed differentiation experiments and analysed the data. G.C. developed and performed the centromere quantification analysis and carried out neuronal differentiation. T.G.-L. assisted in tissue culture and performed karyotype analyses and tissue sectioning. I.S., M.P., U.W. and O.Y. were involved in the genetic screening. M.P. and U.W. assisted with teratoma assays. L.S. assisted with pancreatic differentiation. M.V.S. was involved in all aspects of oocyte donation and research. D.E. derived human pES cell lines from haploid oocytes. D.E. and N.B. supervised the study.

Author Information All high-throughput data have been deposited at the Gene Expression Omnibus (GEO) under accession number GSE71458. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E. (de2220@cumc.columbia.edu) or N.B. (nissimb@cc.huji.ac.il).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Human oocyte manipulation and parthenogenetic ES cell line derivation. Human oocyte donation and pES and swaPS cell line derivation procedures were described previously^{11,23}. Oocyte donors gave informed consent. Experiments were approved by the embryonic stem cell research oversight committee and the institutional review board at Columbia University Medical Center. Briefly, mature MII oocytes were activated using a calcium ionophore and/or an electrical pulse, followed by 4 h of culture with puromycin. Polar body extrusion and the presence of a single pronucleus indicating haploidy were confirmed, and oocytes were allowed to develop to the blastocyst stage. swaPS cells were derived following activation of an oocyte whose nuclear genome had been swapped with that of another oocyte¹¹. ES cell lines were derived by laser ablation of the trophectoderm²⁴ and addition of ROCK inhibitor Y-27632 at 10 μ M to the derivation medium²³. Then 2–3 days after plating, remaining trophectoderm cells were laser ablated, and inner cell mass cells were allowed to grow for 10–14 days until manual picking of the outgrowth was feasible.

Cell culture. Unless otherwise stated, human ES cells were cultured on a feeder layer of growth-arrested mouse embryonic fibroblasts (MEFs) in standard human ES cell medium composed of Knockout Dulbecco's Modified Eagle's Medium supplemented with 15% Knockout Serum Replacement (KSR, Thermo Fisher Scientific), 2 mM L-glutamine, 0.1 mM nonessential amino acids, 50 units ml⁻¹ penicillin, 50 μ g ml⁻¹ streptomycin, 0.1 mM β -mercaptoethanol and 8 ng ml⁻¹ basic fibroblast growth factor (bFGF). Cells were free of mycoplasma and maintained in a humidified incubator at 37°C and 5% CO₂. Passaging was carried out either mechanically with gentle trypsinization using trypsin solution A without EDTA (Biological Industries), or enzymatically using TrypLE Express (Thermo Fisher Scientific) with addition of 10 μ M ROCK inhibitor Y-27632 (Stemgent) for 1 day after splitting. Haploid ES cells could also be grown in feeder-free conditions on Matrigel-coated plates (Corning) in mTeSR1 (STEMCELL Technologies) or StemFitN.AK03 (Ajinomoto) media.

Isolation and maintenance of haploid human ES cell lines. Following identification of haploid cells in human parthenogenetic ES cell lines at passages 6–7 by either metaphase spread analysis or sub-2c-cell sorting (see below and Extended Data Tables 1 and 2), haploid ES cell lines were established by sorting the 1c-cell population, with diploid cells serving as a reference. Haploid ES cell cultures were further maintained by enrichment rounds of 1c-cell sorting every 3–4 passages.

Metaphase spread analysis. For induction of mitotic arrest, growing cells were incubated for 40 min in the presence of 100 ng ml⁻¹ colcemid (Biological Industries), added directly to the culture medium in a humidified incubator at 37°C with 5% CO₂. The cells were then trypsinized, centrifuged at 1,000 r.p.m. at room temperature and gently resuspended in 37°C warmed hypotonic solution (2.8 mg ml⁻¹ KCl and 2.5 mg ml⁻¹ sodium citrate) followed by 20 min of incubation at 37°C. Cells were fixed by addition of fixative solution (3:1 methanol:acetic acid) and incubation for 5 min at room temperature. Fixation was repeated at least three times following centrifugation and resuspension in fixative solution. Metaphase spreads were prepared on slides and stained using the standard G-banding technique. Karyotype integrity was determined according to the International System for Human Cytogenetic Nomenclature (ISCN) based on the observation of a normal karyotype in at least 80% of analysed metaphases (minimum of 20 metaphases per analysis).

Live ES cell sorting by DNA content. Cells were washed with phosphate buffered saline (PBS), dissociated using either TrypLE Select or TrypLE Express (Thermo Fisher Scientific) and stained with 10 μ g ml⁻¹ Hoechst 33342 (ref. 2) (Sigma-Aldrich) in human ES cell medium at 37°C for 30 min. Following centrifugation, cells were resuspended in PBS containing 15% KSR and 10 μ M ROCK inhibitor Y-27632, filtered through a 70- μ m cell strainer (Corning) and sorted using the 405 nm laser in either BD FACSAria III or BD Influx (BD Biosciences). For continued growth, sorted cells were plated with fresh medium containing 10 μ M ROCK inhibitor Y-27632 for 24 h. For comparative analyses, G1-phase cells were sorted from isogenic haploid-enriched and unsorted diploid cultures. Cells that had undergone diploidization relatively recently in culture (within 3 passages after haploid cell enrichment) were isolated by sorting the 4c peak in haploid-enriched cultures and compared with 4c diploid cells from unsorted diploid cultures. Note that haploid-enriched cultures also consist of a mixed 2c-cell population of G2/M-phase haploids and G1-phase diploids. Sorting purity was confirmed by rerunning a fraction of sorted samples through the instrument.

Flow cytometry. All DNA content profiles were generated based on flow cytometry with Hoechst 33342 staining. Haploid cell proportion was estimated based

on the percentage of 1c cells and the relative contribution of G1 cells with regards to other phases of the cell cycle. Estimation of diploidization rate was based on the proportion of haploid cells between consecutive enrichment rounds as well as experimental analysis of h-pES10 diploidization kinetics throughout 7 passages (30 days) by analysing the DNA content of 2–3 replicates at each passage using flow cytometry with propidium iodide in methanol-fixed and RNase-treated cells. Diploidization rate was estimated by fitting the data to an exponential decay curve. For simultaneous flow cytometry analysis of DNA content and cell surface molecules, cells were washed, dissociated and incubated on ice for 30 min in the presence of 10 μ g ml⁻¹ Hoechst 33342 (Sigma-Aldrich) and either a conjugated antibody or a secondary antibody diluted 1:200 following a 60 min incubation with a primary antibody. For simultaneous flow cytometry analysis of DNA content and intracellular PDX1, dissociated cells were treated as described for immunofluorescence procedures, with Hoechst 33342 for DNA staining. Primary antibodies are detailed in Supplementary Table 1. In all flow cytometry procedures, samples were filtered through a 70- μ m cell strainer (Corning Life Sciences) and analysed with either BD FACSAria III or BD Influx (BD Biosciences).

DNA fluorescence *in situ* hybridization. DNA FISH was performed as described elsewhere²⁵ using probes for human chromosomes 2 and 4 and DNA staining with 4',6-diamidino-2-phenylindole (DAPI). Haploidy and diploidy were respectively determined per nucleus based on single or double hybridization signals. ES cells subject to FISH were grown on Matrigel-coated plates in StemFitN.AK03 medium for several passages before analysis.

Alkaline phosphatase and immunofluorescence staining. Alkaline phosphatase staining was performed using the Leukocyte Alkaline Phosphatase Kit (Sigma-Aldrich). For immunofluorescence staining, samples were washed with PBS, fixed with 4% paraformaldehyde for 10 min, and permeabilized and blocked in blocking solution (0.1% Triton X-100 and 5% donkey serum in PBS). Cells were incubated with primary antibodies (detailed in Supplementary Table 1) and secondary antibodies diluted 1:500 in blocking solution, and DAPI was used for DNA staining. Cells were washed twice with PBS subsequently to fixation and each incubation step. Images were taken using Zeiss LSM 510 Meta Confocal Microscope. Centromere quantification was carried out by manually counting centromere foci across individual planes along the z axis. EdU staining was performed using the Click-iT EdU Alexa Fluor 488 Imaging Kit (Thermo Fisher Scientific). ES cells subject to centromere staining in Fig. 1e and Extended Data Fig. 1e were grown on Matrigel-coated plates in StemFitN.AK03 for several passages before analysis.

6-TG resistance screen. To generate a gene trap mutant library, 9 replicates of approximately 4×10^6 haploid pES10 cells (within one passage after 1c-cell enrichment) were co-transfected with 20 μ g 5'-PTK-3' gene trap vector²⁶ and 20 μ g pCyl43 piggyBac transposase plasmid²⁷ using Bio-Rad Gene Pulser (suspended in 800 μ l Opti-MEM, 4-mm cuvettes, 320 V, 250 μ F), and replated on a 100 \times 20 mm dish with DR3 MEFs and ROCK inhibitor Y-27632. Selection for insertions into expressed loci was carried out using 0.3 μ g ml⁻¹ puromycin starting 48 h post transfection, followed by pooling into a single library, represented by approximately 16,000 resistant colonies. Transfection with 5'-PTK-3' only was used as a negative control. To screen for 6-TG-resistant mutants, the mutant library was grown in the presence of 6 μ M 6-TG (Sigma-Aldrich) on DR4 MEFs for 18 days, during which 6 resistant colonies were independently isolated and characterized. Analysis of a resistant clone showed persistence of haploid cells. Genomic DNA was extracted (NucleoSpin Tissue Kit, MACHERY-NAGEL) and insertion sites were detected using splinkerette PCR as described previously²⁸, followed by PCR product purification and Sanger sequencing (ABI PRISM 3730xl DNA Analyzer (Applied Biosystems)). Sequences were mapped to the human genome (GRCh38/hg38) using UCSC BLAT search tool.

Isolation of total DNA and RNA. Total DNA was isolated using the NucleoSpin Tissue Kit (MACHERY-NAGEL). Total RNA was isolated using Qiagen RNeasy Kits according to the manufacturer's protocols. To determine total RNA levels per cell, haploid and diploid cells were isolated from the same cultures by sorting the 1c (haploid G1) and 4c (diploid G2/M) populations, respectively. Following growth for 2 passages, cells were harvested and counted, and RNA was isolated from triplicates of 400,000 cells from each cell line and ploidy state (pES10 and pES12, haploid and diploid; 12 samples in total). RNA amounts were quantified using NanoDrop. **Genome integrity analysis.** Copy number variation (CNV) analysis was carried out on DNA samples of G1-sorted haploid and diploid pES10 and pES12 cells (see Supplementary Table 2) using Infinium Omni2.5Exome-8 BeadChip single nucleotide polymorphism (SNP) arrays (Illumina) following the manufacturer's protocols. Raw data were processed using Genome Studio Genotyping Module (Illumina) to obtain log R ratios values for analysis using R statistical programming language. As expected, diploid pES10 and pES12 cells were homozygous across all chromosomes.

RNA sequencing. For a detailed list of samples analysed by RNA-seq, see Supplementary Table 3. Total RNA samples (200 ng–1 µg, RNA integrity number (RIN) >9) were enriched for mRNAs by pulldown of poly(A)⁺ RNA. RNA-seq libraries were prepared using the TruSeq RNA Library Prep Kit v2 (Illumina) according to the manufacturer's protocol and sequenced using Illumina NextSeq 500 to generate 85 bp single-end reads.

Transcriptome analysis. RNA-seq reads were aligned to the human reference genome (GRCh37/hg19) using TopHat (version 2.0.8b) allowing 5 mismatches. Reads per kilobase per million fragments mapped (RPKM) values were quantified using Cuffquant and normalized using Cuffnorm in Cufflinks (version 2.1.1) to generate relative gene expression levels. Hierarchical clustering analyses were performed on RPKM values using Pearson correlation and average linkage. Analysis of differential gene expression relative to total RNA in haploid and diploid human ES cells ($n = 4$ in each group) was carried out by two complementary strategies, as follows: first, we used Cuffdiff with default parameters, considering differences of greater than twofold with FDR <0.05 as significant; second, to identify possibly subtle yet consistent transcriptional differences, we tested for genes whose minimal expression levels across all replicates of a certain group were higher than their maximal expression level across all replicates of the other group. Statistical significance was then determined by two-tailed unpaired Student's *t*-test. Functional annotation enrichment analysis was done by DAVID (using the Benjamini method to determine statistical significance). Imprinting analyses included 75 human imprinted genes (<http://www.geneimprint.com/>), listed in Supplementary Table 4. RNA-seq data from control ES cell line NYSCF1 were published elsewhere²⁹ (GEO accession number GSE61657). Genome-wide gene expression moving median plots were generated using the R package zoo (version 1.7–12) after removal of genes that were not expressed in the averaged reference diploid sample by flooring to 1 and setting an expression threshold of above 1. RNA-seq data from different tissues were retrieved from the Genotype-Tissue Expression (GTEx) portal (<http://www.gtexportal.org/>)³⁰. Colour-coded scales in Fig. 4d correspond to gene expression levels relative to the mean across tissues (left scale) and across each set of ES cell duplicate and EB sample (right scale). Expression microarray analysis was performed as previously³¹ by using Affymetrix Human Gene 1.0 ST arrays.

DNA methylation analysis. DNA methylation analysis was performed on genomic DNA from the samples detailed in Supplementary Table 2 using Infinium HumanMethylation450 BeadChips (Illumina) following the Infinium HD Methylation Protocol as described previously²⁹. DNA methylation data from control ES cell line NYSCF1 were published before²⁹ (GEO accession number GSE61657). Data were processed and normalized by using subset-quantile within array normalization (SWAN) and adjusted for batch effects using the R package ChAMP (version 1.4.0). DNA methylation levels at CpG sites associated with pluripotency-specific genes and iDMRs were analysed as described before²⁹. For analysis of DNA methylation levels on the X chromosome, probes with average β values of less than 0.4 were filtered out. DMR analysis was facilitated by the lasso function in ChAMP using default settings. DMRs were then assigned to genes by proximity and analysed for functional annotation enrichment using DAVID (using the Benjamini method to determine statistical significance).

Cell size analysis. Following sorting of haploid and diploid cell populations in G1, the diameter ($2r$) of viable single cells was measured by Countess Automated Cell Counter (Invitrogen) and their surface area and volume were calculated as $4\pi r^2$ and $4/3\pi r^3$, respectively. Analysis included 7, 4, 8 and 4 technical replicates for 1n pES10, 1n pES12, 2n pES10 and 2n pES12, respectively.

Mitochondrial DNA abundance analysis. Relative mtDNA abundance was analysed by quantitative PCR (qPCR) by using primers for the mitochondrial gene *MT-ND2* (forward primer: 5'-TGTTGGTTATACCCCTTCCCGTACTA-3'; reverse primer: 5'-CCTGCAAAGATGGTAGAGTAGATGA-3') and normalization to nuclear DNA by using primers for the nuclear gene *BECN1* (forward primer: 5'-CCCTCATCACAGGGCTCTCTCCA-3'; reverse primer: 5'-GGGACTGTAGGTGGGAAGTATGC-3'), as described elsewhere³². Analysis was performed using Applied Biosystems 7300 Real-Time PCR System with PerfeCTa SYBR Green FastMix (Quanta Biosciences). Analysis included all G1-sorted samples detailed in Supplementary Table 2 ($n = 4$ for each group, with two biological replicates for each cell line).

Embryoid body differentiation. EB differentiation was carried out by detaching ES cell colonies with Trypsin solution A without EDTA (Biological Industries), followed by resuspension and further culture of cell aggregates in human ES cell medium without bFGF on low attachment plates. Differentiation of haploid ES cells was initiated within 2 passages after 1c-cell enrichment. After 21 days, EB RNA was extracted from unsorted and/or sorted EB cells in G1 following dissociation and staining with 10 µg ml⁻¹ Hoechst 33342 (Sigma-Aldrich) at 37 °C for 30 min.

Metaphase spread analysis was performed on dissociated EB cells plated on 0.2% gelatin and expanded in human ES cell medium without bFGF.

Differentiation into neural progenitor cells. NCAM1-positive ES cell-derived neural progenitor cells were obtained using a 10-days protocol for efficient neural differentiation³³ with slight modification³⁴. Differentiation was initiated within 2 passages after 1c-cell enrichment. RNA was extracted from sorted haploid NCAM1-positive cells in G1 by co-staining with Hoechst 33342 and an anti-human NCAM-1/CD56 primary antibody (see Supplementary Table 1) and a Cy3-conjugated secondary antibody (Jackson Immunoresearch Laboratories) diluted 1:200.

Neuronal differentiation. Differentiation into neurons was carried out by following a published protocol³⁵ based on synergistic inhibition of SMAD signalling³⁶ with modification, as follows: differentiation was initiated within 2 passages after 1c-cell enrichment with fully confluent ES cells cultured on Matrigel-coated plates in mTeSR1 by replacing the medium with human ES cell medium without bFGF, containing 10 µM SB431542 (Selleckchem) and 2.5 µM LDN-193189 (Stemgent) for 4 days. Subsequently, cells were kept in N2 medium³⁵ supplemented with 10 µM SB431542 and 2.5 µM LDN-193189 for an additional 4 days, followed by 2 days in N2 medium supplemented with B-27 (Thermo Fisher Scientific) and 10 µM DAPT (Stemgent). The cells were then dissociated and replated on 0.01% poly-L-ornithine coated (Sigma-Aldrich) and laminin coated (4 µg ml⁻¹, Thermo Fisher Scientific) plates in the presence of 10 µM ROCK inhibitor Y-27632 (Selleckchem), and further cultured in the same medium without Y-27632 for the next 4 days. Neuronal cultures were maintained in N2 medium supplemented with B-27 and 20 ng ml⁻¹ BDNF (R&D) until analysis by immunostaining and FISH on day 20.

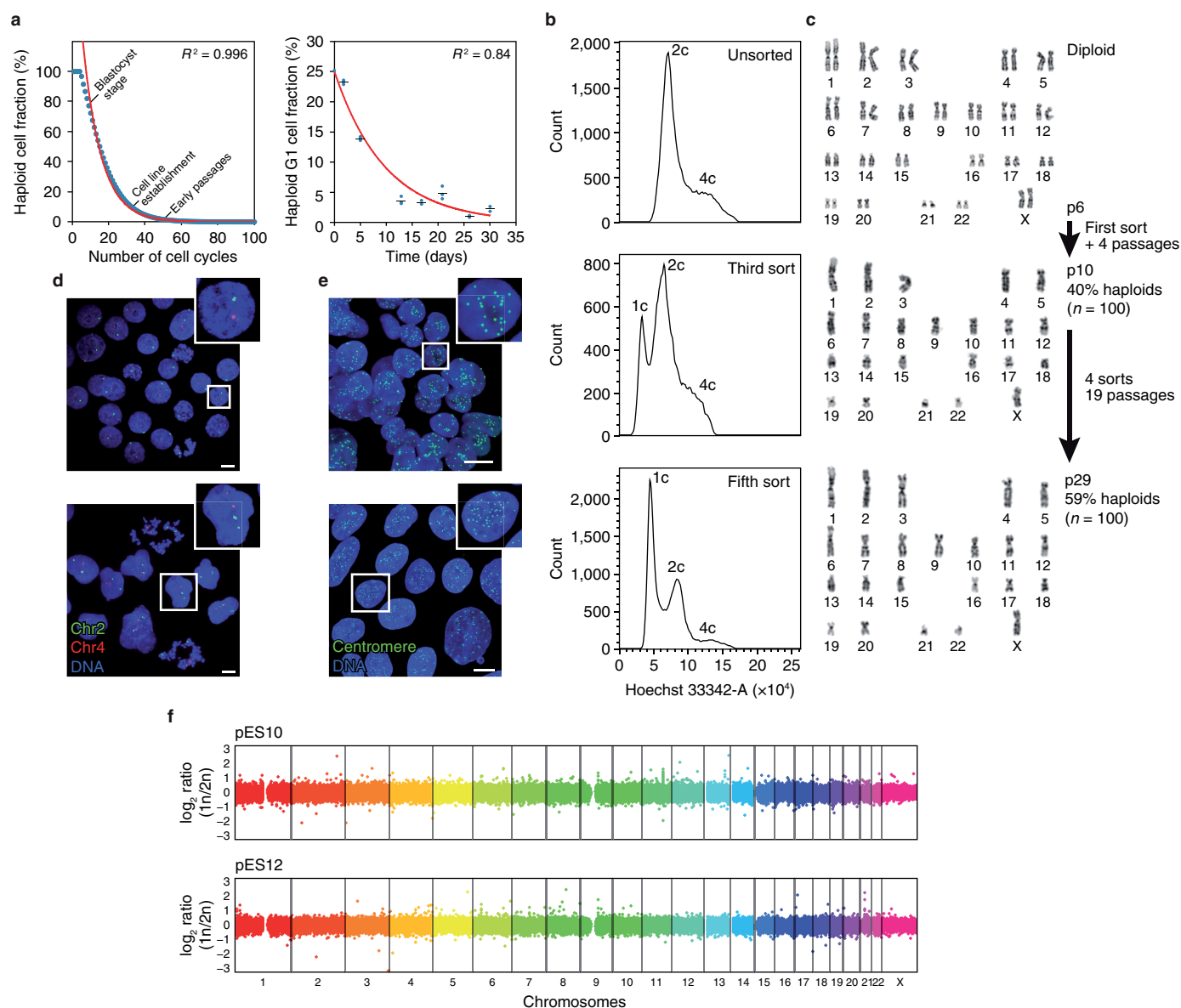
Cardiomyocyte differentiation. 80–90% confluent ES cells grown on Matrigel-coated plates in mTeSR1 were subject to an 11-day regimen³⁷ based on consecutive GSK3 and WNT inhibition with CHIR99021 and IWP-2 (Selleckchem), respectively. Differentiation was initiated within 2 passages after 1c-cell enrichment. On day 11 of differentiation, 1c cells were sorted and plated for immunostaining.

Differentiation towards the pancreatic lineage. The protocol used here was developed based on several recent publications^{38–40}. ES cells grown in feeder-free conditions were differentiated into definitive endoderm by using STEMdiff Definitive Endoderm Kit (StemCell Technologies) for 3–4 days. Subsequent specification was achieved by a step-wise protocol involving treatment with recombinant human KGF/FGF7 (R&D Systems), LDN-193189 (Stemgent), KAAD-cyclopamine (Stemgent) and retinoic acid (Stemgent). On days 8–11, EGF (R&D System) was used to induce pancreatic cells. Differentiation was initiated within as few as 2 passages after 1c-cell enrichment.

Teratoma formation assay. All experimental procedures in animals were approved by the ethics committee of the Hebrew University. ES cells were trypsinized and approximately 2×10^6 cells were resuspended in 100 µl human ES cell medium and 100 µl Matrigel (BD Biosciences), followed by subcutaneous injection into NOD-SCID *Il2rg*^{-/-} immunodeficient mice (Jackson Laboratory). 8–12 weeks after injection tumours were dissected and subjected to further analysis. Histological slides were prepared from tumour slices cryopreserved in O.C.T. compound (Sakura Finetek) using Leica CM1850 cryostat (Leica Biosystems, 10-µm sections), followed by immunostaining, haematoxylin and eosin staining or FISH analysis. Flow cytometry with Hoechst 33342 staining was performed on dissociated cells from freshly dissected tumours.

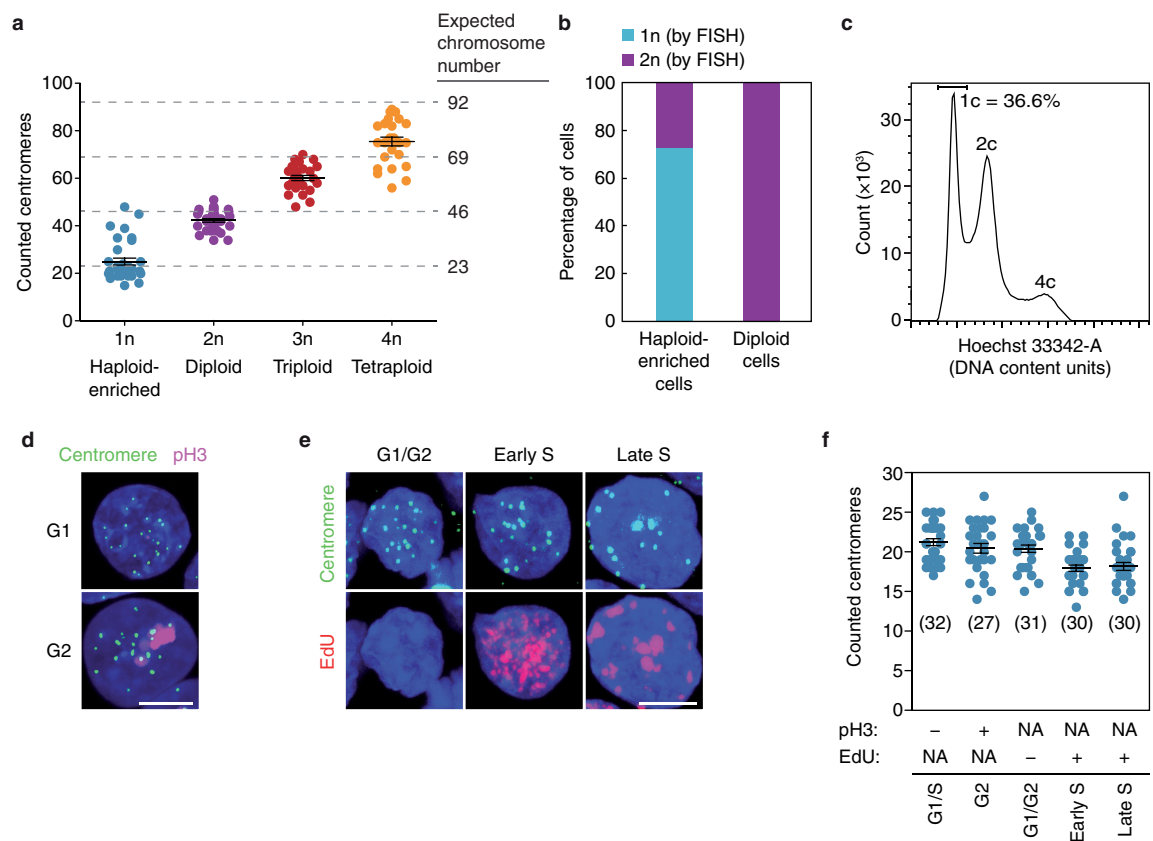
23. Noggle, S. *et al.* Human oocytes reprogram somatic cells to a pluripotent state. *Nature* **478**, 70–75 (2011).
24. Chen, A. E. *et al.* Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* **4**, 103–106 (2009).
25. Rao, P. H., Nandula, S. V. & Murty, V. V. Molecular cytogenetic applications in analysis of the cancer genome. *Methods Mol. Biol.* **383**, 165–185 (2007).
26. Cadiñanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* **35**, e87 (2007).
27. Wang, W. *et al.* Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 9290–9295 (2008).
28. Chen, L. *et al.* Transposon activation mutagenesis as a screening tool for identifying resistance to cancer therapeutics. *BMC Cancer* **13**, 93 (2013).
29. Johansson, B. *et al.* Comparable frequencies of coding mutations and loss of imprinting in human pluripotent cells derived by nuclear transfer and defined factors. *Cell Stem Cell* **15**, 634–642 (2014).
30. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature Genet.* **45**, 580–585 (2013).
31. Yamada, M. *et al.* Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells. *Nature* **510**, 533–536 (2014).
32. Wanet, A. *et al.* Mitochondrial remodeling in hepatic differentiation and dedifferentiation. *Int. J. Biochem. Cell Biol.* **54**, 174–185 (2014).
33. Kim, D.-S. *et al.* Robust enhancement of neural differentiation from human ES and iPS cells regardless of their innate difference in differentiation propensity. *Stem Cell Rev.* **6**, 270–281 (2010).

34. Stelzer, Y., Sagi, I. & Benvenisty, N. Involvement of parental imprinting in the antisense regulation of onco-miR-372-373. *Nat. Commun.* **4**, 2724 (2013).
35. Wang, L. *et al.* Differentiation of hypothalamic-like neurons from human pluripotent stem cells. *J. Clin. Invest.* **125**, 796–808 (2015).
36. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnol.* **27**, 275–280 (2009).
37. Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ β -catenin signaling under fully defined conditions. *Nature Protocols* **8**, 162–175 (2013).
38. Hua, H. *et al.* iPSC-derived β cells model diabetes due to glucokinase deficiency. *J. Clin. Invest.* **123**, 3146–3153 (2013).
39. Pagliuca, F. W. *et al.* Generation of functional human pancreatic β cells *in vitro*. *Cell* **159**, 428–439 (2014).
40. Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived *in vitro* from human pluripotent stem cells. *Nature Biotechnol.* **32**, 1121–1133 (2014).
41. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).



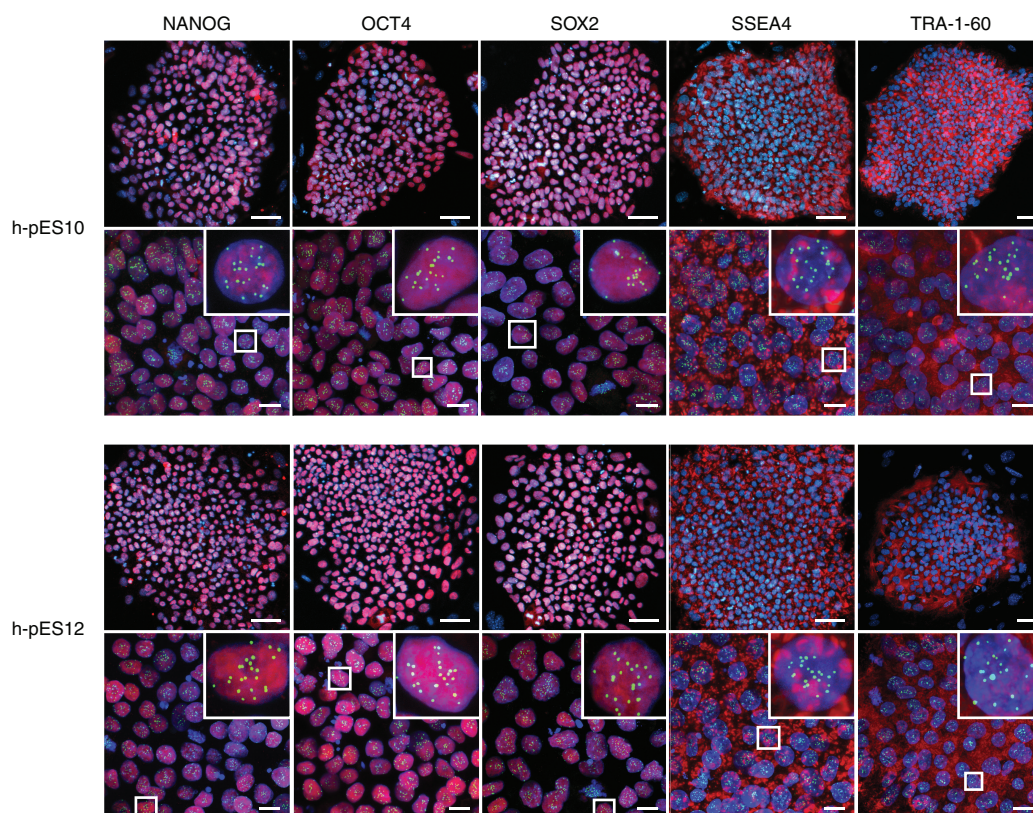
Extended Data Figure 1 | Derivation of haploid human ES cell lines h-pES10 and h-pES12. **a**, Left panel, diploidization rate model for a haploid egg with a theoretical diploidization probability of 10%, overlaid with an exponential decay fit (red curve, see Supplementary Notes). Approximated cell cycle numbers for different ES cell line derivation stages are indicated. Right panel, diploidization dynamics of h-pES10 over seven passages by flow cytometry, overlaid with an exponential fit to the data (red curve). Bars indicate the mean across biological replicates. **b**, Establishment of a haploid-enriched human ES cell line from pES12 cells after repeated sorting and enrichment of 1c cells using Hoechst 33342 staining. Top to bottom, DNA content profiles of

unsorted diploid cells, partially purified haploid cells at the third sort, and mostly-purified haploid cells at the fifth sort. **c**, Karyotypes and haploid metaphase percentage over the course of enrichment and passaging. **d, e**, DNA FISH (**d**) and centromere protein immunofluorescence staining (**e**) in haploid-enriched pES12 cells (upper panels) and unsorted diploid pES10 cells (lower panels). Magnified insets show representative haploid and diploid nuclei with single or double hybridization signals (**d**) and 23 or 46 centromeres (**e**), respectively. Scale bars, 10 μm . **f**, SNP array-based CNV analysis comparing haploid pES10 and pES12 cells with their unsorted diploid counterparts (logarithmic scale).



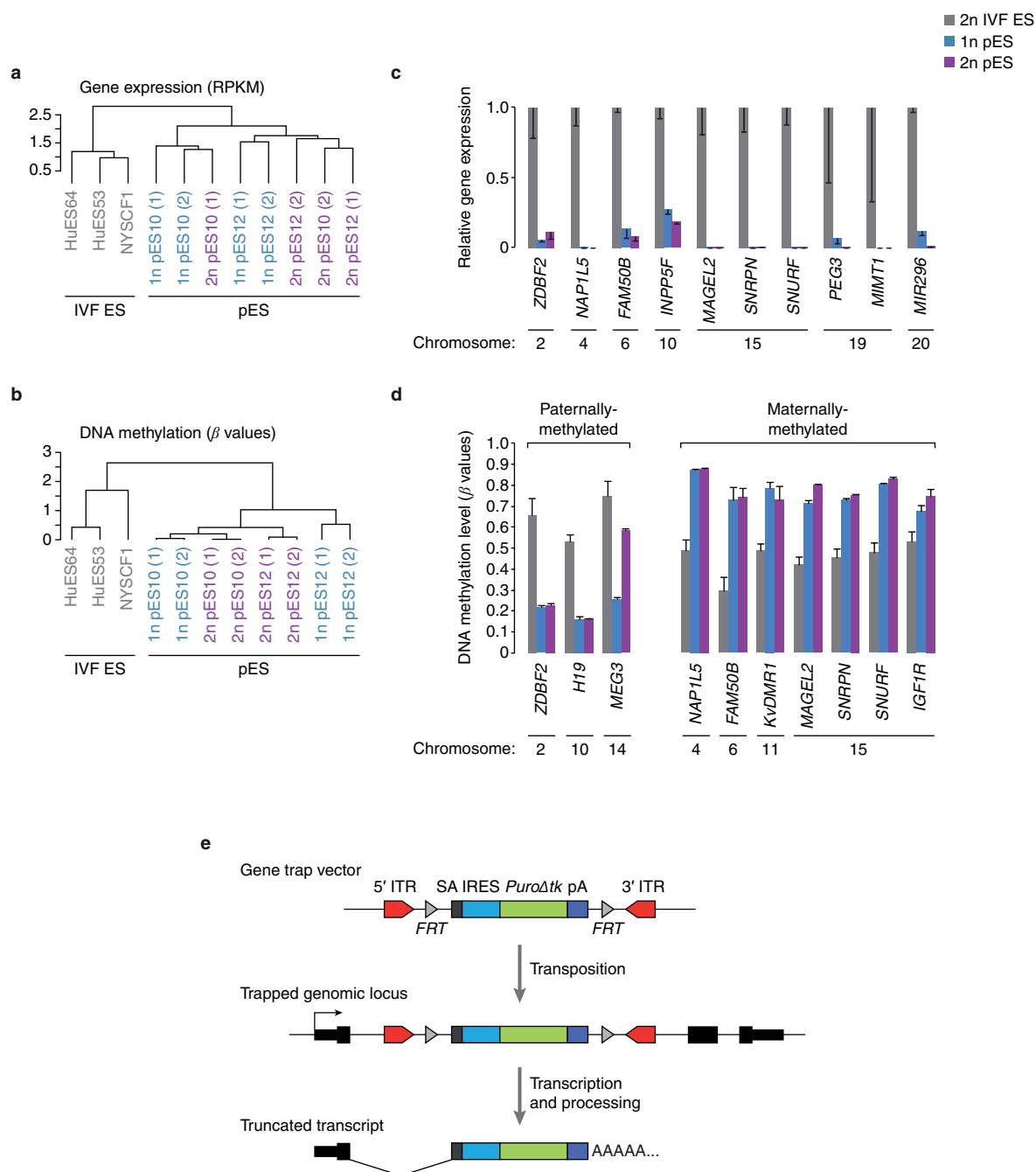
Extended Data Figure 2 | Determination of ploidy at single-cell level by quantification of centromere foci. **a**, The counted number of centromeres correlates with ploidy. 1n, haploid-enriched pES10 cells grown for 4 passages after the fourth sort ($n = 33$; 76% haploids by this assay); 2n, unsorted diploid pES10 cells ($n = 34$); 3n, soPS2 cells²³ ($n = 27$); 4n, Hybrid1 cells⁴¹ ($n = 27$). Black horizontal lines indicate mean \pm s.e.m. and dashed lines mark expected chromosome numbers. **b**, Quantification of haploid and diploid cells by DNA FISH in the haploid-enriched ($n = 152$; 73% haploids by this assay) and diploid ($n = 135$) cells in **a**. **c**, DNA

content profile of the haploid-enriched cells in **a** (73% haploids by this assay). **c**, chromosomal copies. **d**, **e**, Co-staining of centromeres and either phospho-histone 3 (pH3, Ser10) (**d**) or 5-ethynyl-2'-deoxyuridine (EdU) (**e**) for distinguishing between different stages of interphase in haploid pES12 cells. In blue, DNA staining. Scale bar, 5 μ m. **f**, Quantification of centromere counts in the different cell cycle stages shown in **d** and **e**. n indicated in parenthesis. Black horizontal lines indicate mean \pm s.e.m. See Supplementary Notes for details.



Extended Data Figure 3 | Pluripotent stem cell markers in haploid human ES cells. Co-staining of pluripotency markers NANOG, OCT4, SOX2, SSEA4 and TRA-1-60 (red), centromeres (green) and DNA (blue)

in h-pES10 and h-pES12 at colony resolution (upper panels; scale bars, 50 μm) and single-cell resolution (lower panels; scale bars, 10 μm). Magnified insets show representative haploid cells with 23 centromeres.

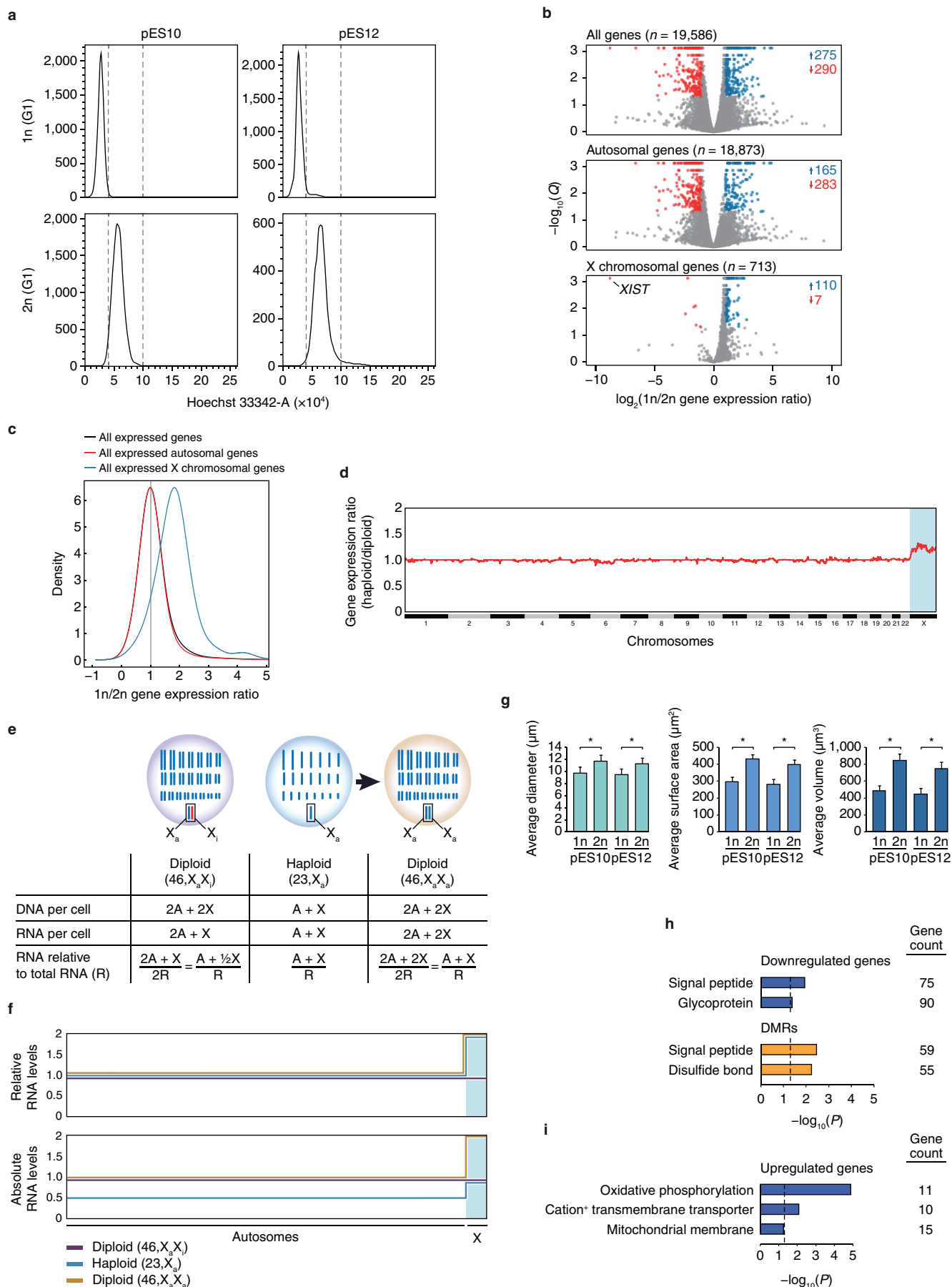


Extended Data Figure 4 | Analysis of parental imprinting and gene trap mutagenesis in haploid human parthenogenetic ES cells.

a, b, Clustering analysis of diploid *in vitro* fertilization (IVF) ES cells and G1-sorted haploid and diploid parthenogenetic ES (pES) cells by expression levels of imprinted genes ($n = 75$, see Supplementary Table 4) (**a**) and DNA methylation levels at imprinted differentially methylated regions (iDMRs, $n = 35$)²⁹ (**b**). The symbols (1) and (2) indicate biological replicates. **c**, Relative mean expression levels \pm s.e.m. of representative paternally expressed imprinted genes across seven chromosomes in the samples shown in **a** (RPKM ratios). **d**, Mean DNA methylation levels \pm s.e.m. at representative paternally methylated and maternally methylated iDMRs (typically intermediately methylated in bi-parental

control cells, and respectively hypomethylated and hypermethylated in parthenogenetic cells) in the samples shown in **b**. β values range from complete hypomethylation (0) to complete hypermethylation (1).

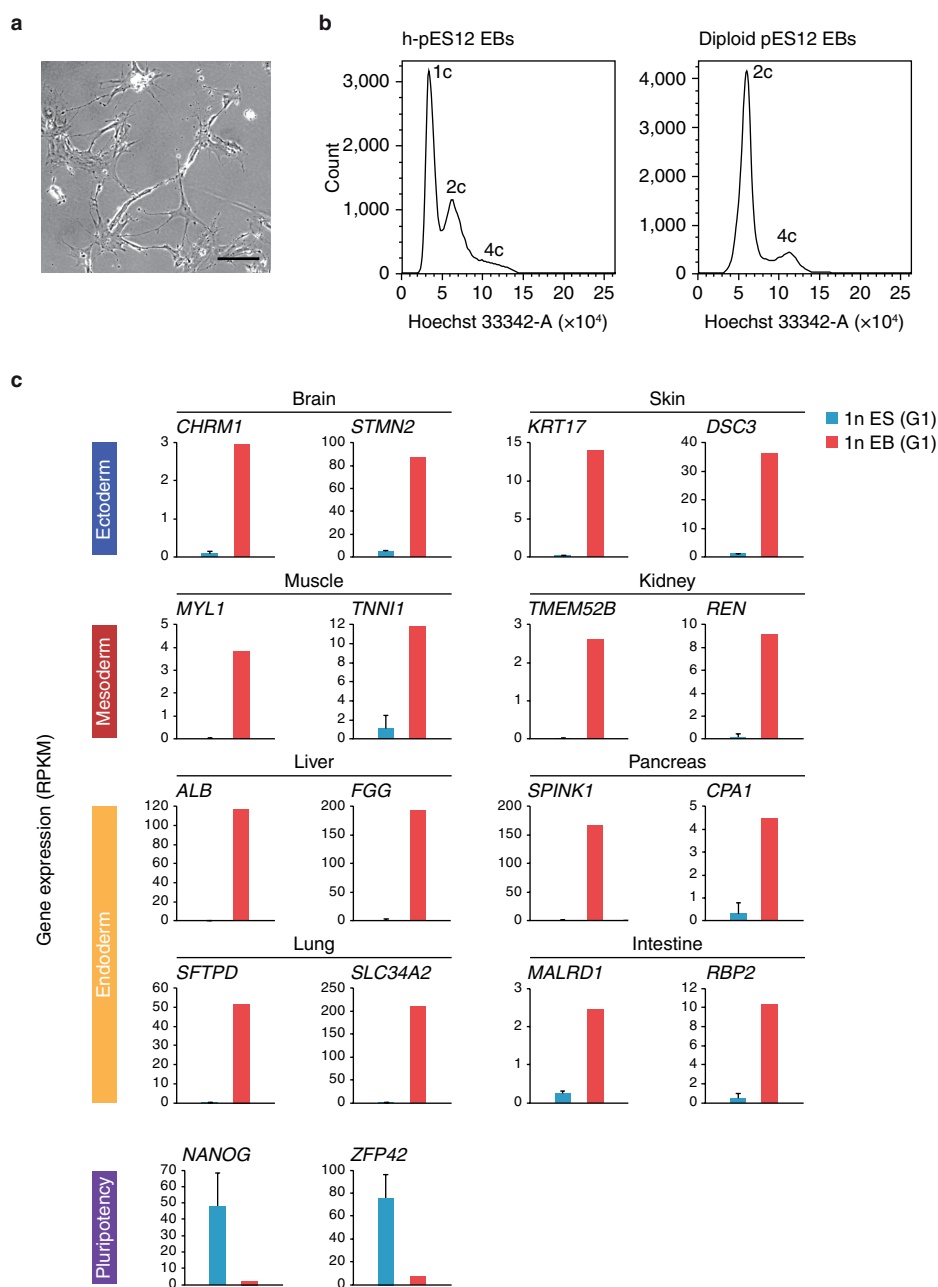
e, Schematic outline of the *piggyBac* gene trap system. The gene trap vector²⁶ is flanked by *piggyBac* inverted terminal repeats (ITRs) and *FRT* sites, and carries a 5' splice acceptor (SA), an internal ribosome entry site (IRES) element followed by a promoterless puromycin resistance gene (*Puro Δ tk*) and a 3' poly(A) signal (pA). In the presence of the *PiggyBac* transposase (encoded on a separate plasmid²⁷, not shown), the gene trap vector undergoes random transposition into the genome. Insertion into a transcriptionally active gene results in truncation of the endogenous transcript and introduction of resistance to puromycin.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Comparative analyses of isogenic haploid and diploid human ES cells. **a**, Sorting purity of haploid and diploid ES cells in G1. **b**, Log-scaled volcano plots of relative differential gene expression between haploid and diploid human ES cells, divided into panels by all genes (top), autosomal genes (middle) and X chromosomal genes (bottom). Q , false discovery rate (FDR). Significantly downregulated and upregulated genes (greater than twofold change, $Q < 0.05$) in haploid cells are marked in red and blue, respectively, and their totals are indicated to the right. Note that *XIST* is the most downregulated transcript in haploid cells. **c**, Smoothed distributions of the 1n/2n gene expression ratios for all expressed genes, all expressed autosomal genes and all expressed X chromosomal genes (expression threshold, mean RPKM > 0.1). **d**, Genome-wide moving median plot of the gene expression ratio between haploid and diploid pES10 cells in G1 by expression microarray analysis (window size = 100 genes). **e**, **f**, Model for genome-wide autosomal

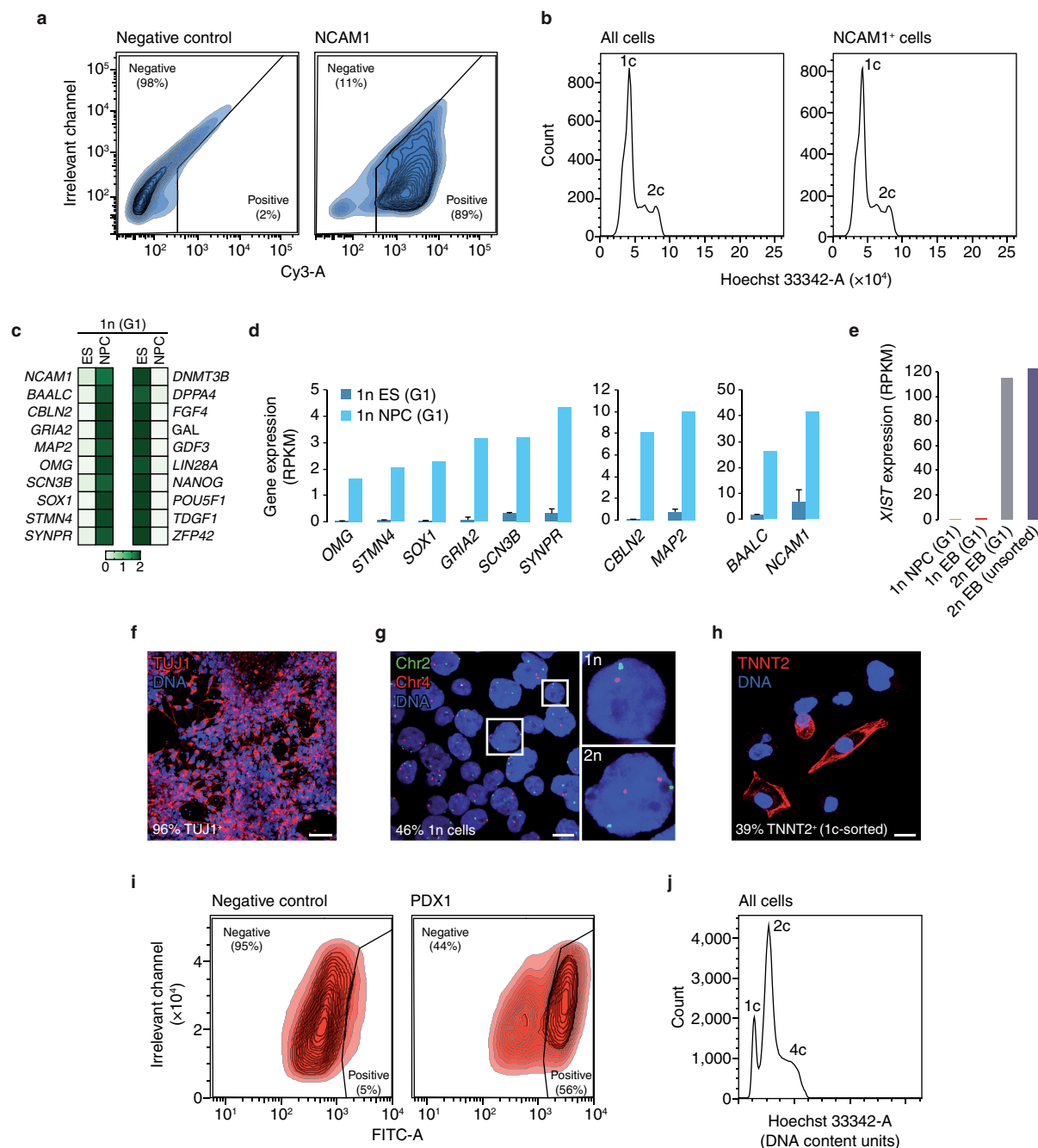
gene expression level reduction in haploid human ES cell as inferred by differential X chromosome inactivation status. **e**, DNA content, RNA expression levels relative to total RNA and presumed equality of absolute X chromosomal gene dosage in haploid (X_a) and diploid (X_aX_i) human ES cells, enable the estimation of total RNA levels per haploid cell. X_a and X_i denote active (blue) and inactive (red) X chromosomes, respectively. A, autosomes; X, X chromosome; R, total RNA. **f**, Schematic genome-wide representation of relative and absolute RNA levels in the cells shown in **e**. **g**, Average diameter and calculated surface area and volume of G1-sorted haploid and diploid ES cells. Error bars represent s.d. $*P < 0.01$ (two-tailed unpaired Student's t -test). **h**, **i**, Functional annotation enrichment analysis for relatively downregulated genes and differentially methylated regions (DMRs) (**h**), as well as relatively upregulated genes (**i**) in haploid ES cells compared with diploid ES cells.



Extended Data Figure 6 | EB differentiation of haploid human ES cells.

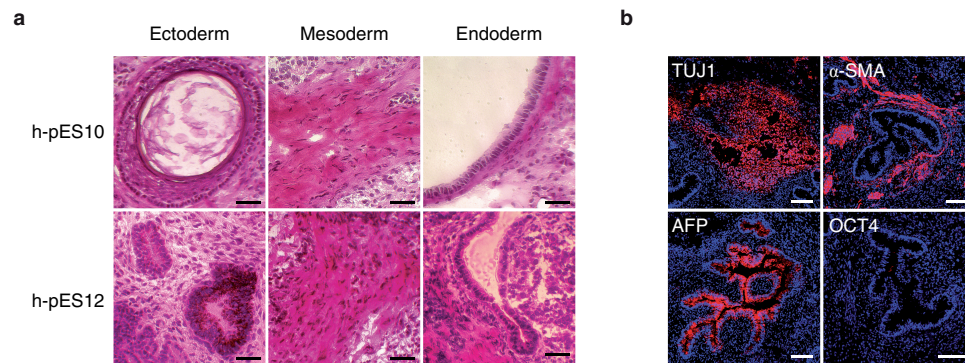
a, Representative image of plated cells dissociated from h-pES12-derived 21-day EBs. Karyotype is shown in Fig. 4b. Scale bar, 100 μ m. **b**, DNA content profiles of dissociated EBs derived from haploid-enriched and

diploid pES12 cells. **c**, Expression levels (RPKM) of tissue- and pluripotency-specific genes in undifferentiated (ES) and differentiated (EB) G1-sorted haploid (1n) pES10 cells.



Extended Data Figure 7 | Directed differentiation of haploid human ES cells. **a, b**, Flow cytometry analysis with co-staining of DNA and NCAM1 in h-pES10 cells following neural differentiation. **a**, Gating for NCAM1-positive cells (right) based on a negative secondary antibody stained control sample (left). **b**, DNA content profiles of the entire cell population (left) and NCAM1-positive cells (right). **c**, chromosomal copies. **c, d**, Expression levels of neural- and pluripotency-specific genes in haploid G1-sorted ES cells and NPCs. **e**, *XIST* expression levels in haploid and diploid pES10-derived EBs and NPCs. **f**, TUJ1 staining in h-pES12-derived neurons. Scale bar, 100 μm. **g**, DNA FISH in the

neurons shown in **f**. Magnified insets show representative haploid and diploid nuclei with single and double hybridization signals, respectively. Scale bar, 10 μm. **h**, TNNT2 staining in G1-sorted haploid pES12-derived cardiomyocytes. Scale bar, 10 μm. **i, j**, Flow cytometry analysis with co-staining of DNA and PDX1 in h-pES10 cells following pancreatic differentiation. **i**, Gating for PDX1-positive cells (right) based on a negative secondary antibody stained control sample (left). **j**, DNA content profile of the entire cell population (related to Fig. 4k). **c**, chromosomal copies.



Extended Data Figure 8 | *In vivo* differentiation of haploid human ES cells. **a**, Haematoxylin and eosin histological sections of teratomas derived from h-pES10 and h-pES12. Scale bar, 50 μ m. **b**, TUJ1 (ectoderm), α -SMA

(mesoderm), AFP (endoderm) and OCT4 (pluripotency) staining in an h-pES10-derived teratoma. DNA staining is shown in blue. Note the absence of nuclear OCT4 staining. Scale bars, 100 μ m.

Extended Data Table 1 | Identification of haploid cells in early-passage human parthenogenetic ES cell lines by metaphase spread analysis

pES cell line	Oocyte donor	Passage number at analysis	Number of haploid metaphases	Number of diploid metaphases	Total number of metaphases	% Haploid metaphases
1	8/31/09	4	0	233	233	0
2	1043	6	0	278	278	0
3	1058	6	0	273	273	0
4	1058	6	0	222	222	0
5	1058	9	0	50	50	0
6	1105	5	0	13	13	0
		10	0	140	140	0
8	1151	4	0	361	361	0
9	1157	6	0	234	234	0
10	1160	7	2	150	152	1.32
11	1160	8	0	194	194	0
Total:			2	2148	2150	

The derivation of cell lines pES1–6 was reported previously^{11,23}.

Extended Data Table 2 | Isolation of haploid cells from early-passage human parthenogenetic ES cell lines by sub-2c-cell sorting

pES cell line	Oocyte donor	Passage number at first sort	Passage number at second sort	Presence of haploid cells at second sort
6	1105	6	10	Undetected
12	1160	6	10	Yes
swaPS cell line*				
4	1126	6	11	Undetected
5	1155	5	9	Undetected
11	1175	4	Sorted population did not survive	

*swaPS cells are parthenogenetic ES cells derived following activation of an oocyte whose nuclear genome had been swapped with that of another oocyte¹¹.

Mitochondrial ROS regulate thermogenic energy expenditure and sulfenylation of UCP1

Edward T. Chouchani^{1,2*}, Lawrence Kazak^{1,2*}, Mark P. Jedrychowski², Gina Z. Lu^{1,2}, Brian K. Erickson², John Szpyt², Kerry A. Pierce³, Dina Laznik-Bogoslavski¹, Ramalingam Vetrivelan⁴, Clary B. Clish³, Alan J. Robinson⁵, Steve P. Gygi² & Bruce M. Spiegelman^{1,2}

Brown and beige adipose tissues can dissipate chemical energy as heat through thermogenic respiration, which requires uncoupling protein 1 (UCP1)^{1,2}. Thermogenesis from these adipocytes can combat obesity and diabetes³, encouraging investigation of factors that control UCP1-dependent respiration *in vivo*. Here we show that acutely activated thermogenesis in brown adipose tissue is defined by a substantial increase in levels of mitochondrial reactive oxygen species (ROS). Remarkably, this process supports *in vivo* thermogenesis, as pharmacological depletion of mitochondrial ROS results in hypothermia upon cold exposure, and inhibits UCP1-dependent increases in whole-body energy expenditure. We further establish that thermogenic ROS alter the redox status of cysteine thiols in brown adipose tissue to drive increased respiration, and that Cys253 of UCP1 is a key target. UCP1 Cys253 is sulfenylated during thermogenesis, while mutation of this site desensitizes the purine-nucleotide-inhibited state of the carrier to adrenergic activation and uncoupling. These studies identify mitochondrial ROS induction in brown adipose tissue as a mechanism that supports UCP1-dependent thermogenesis and whole-body energy expenditure, which opens the way to improved therapeutic strategies for combating metabolic disorders.

The specialized capacity for thermogenic respiration in brown adipose tissue (BAT) and beige fat⁴ relies on UCP1. This mitochondrial inner membrane carrier dissipates protonmotive force (Δp) and increases the rate of substrate oxidation to generate heat¹. While UCP1 is critical for thermogenesis, mechanisms that support UCP1-dependent respiration *in vivo* are not fully understood^{1,2}.

We examined features of mitochondrial function in mouse interscapular BAT after cold (4 °C) exposure. Unexpectedly, we found that acute activation of BAT thermogenesis *in vivo* was associated with a substantial increase in mitochondrial ROS levels. Mitochondrial superoxide (Fig. 1a), lipid hydroperoxides (Fig. 1b), and mitochondrial hydrogen peroxide (Fig. 1c)⁵ levels increased significantly in BAT after acute cold exposure. Furthermore, primary brown adipocytes exposed to adrenergic stimulation of thermogenesis by noradrenaline exhibited acute dose-dependent increases in superoxide levels (Extended Data Fig. 1a). Having identified the induction of mitochondrial ROS as an early event during thermogenesis in BAT, we investigated whether it was a regulator of function *in vivo*. We injected mice intraperitoneally (i.p.) with the mitochondria-targeted antioxidant MitoQ before cold exposure. MitoQ efficiently depletes mitochondrial free radical species *in vivo* and achieves substantial localization in metabolically active tissues including adipose⁶. Treatment with MitoQ before cold exposure inhibited the cold-dependent increases in BAT mitochondrial superoxide and lipid hydroperoxides (Fig. 1a, b). Notably, MitoQ administration resulted in a dose-dependent induction of hypothermia upon cold exposure (Fig. 1d). MitoQ had no effect on

body temperature of mice without cold stress (Fig. 1d), and equimolar injections of decylTTP, the mitochondria-targeting moiety lacking the antioxidant group, had no significant effect on body temperature during cold exposure (Extended Data Fig. 1b).

While BAT and beige fat are the primary tissues for mediating adaptive thermogenesis, muscle shivering also plays an important role, particularly during the early thermogenic response⁷. Shivering was monitored directly by electromyography (EMG) after the transition from thermoneutrality to cold. As expected, by 30 min cold exposure substantial shivering was observed (Extended Data Fig. 1c–e), which could be abrogated by the nicotinic acetylcholine receptor inhibitor curare (Extended Data Fig. 1c). Importantly, treatment of mice with MitoQ before cold exposure caused no detectable change in shivering activity (Extended Data Fig. 1c–e).

We next tested whether the hypothermic effects of depleting mitochondrial ROS depended on UCP1. Mice genetically lacking UCP1 (UCP1^{−/−}) are sensitive to acute cold challenge, but maintain thermal homeostasis after gradual acclimation to temperatures below thermoneutrality through compensatory upregulation of alternative thermogenic pathways⁸. We injected wild-type (WT) and UCP1^{−/−} mice with MitoQ after gradual cold acclimation. Notably, UCP1^{−/−} mice were protected against the hypothermic effects of MitoQ (Fig. 1e). Moreover, mice injected with MitoQ 10 min before β 3-adrenergic stimulation by CL 316,243 (CL)⁹ exhibited reduced CL-dependent oxygen consumption (Fig. 1f, g and Extended Data Fig. 1f). These differences were observed in the absence of any significant changes in physical movement (Extended Data Fig. 1g). Together, these data provide evidence that mitochondrial ROS support UCP1-dependent thermogenesis *in vivo*.

It is increasingly appreciated that mitochondrial ROS act as signalling molecules in physiology, which involves oxidative modification of sensitive cysteine thiol residues on proximal protein targets¹⁰. Strikingly, acute cold exposure drove substantial oxidation and depletion of the BAT glutathione pool (Extended Data Fig. 2a). Furthermore, by performing quantitative profiling of *in vivo* protein thiol redox status in BAT using ratiometric labelling with iodoTMT (Extended Data Fig. 2b), we found that BAT protein thiols became substantially more oxidized upon cold exposure (Fig. 2a and Extended Data Fig. 2c). Pathway analysis of those protein cysteine targets most substantially oxidized during thermogenesis revealed pathways involved in mitochondrial metabolism to be the most enriched (Fig. 2b and Supplementary Table 1). Finally, we profiled levels of protein sulfenic acids¹¹—the proximal reversible cysteine modification generated by lipid hydroperoxides and hydrogen peroxide¹²—and found them to be significantly elevated upon cold exposure (Fig. 2c). These findings suggested that oxidation of mitochondrial thiols might be important for signalling ROS-dependent thermogenesis in BAT.

¹Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ⁴Department of Neurology, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁵MRC Mitochondrial Biology Unit, Hills Road, Cambridge CB2 0XY, UK.

*These authors contributed equally to this work.

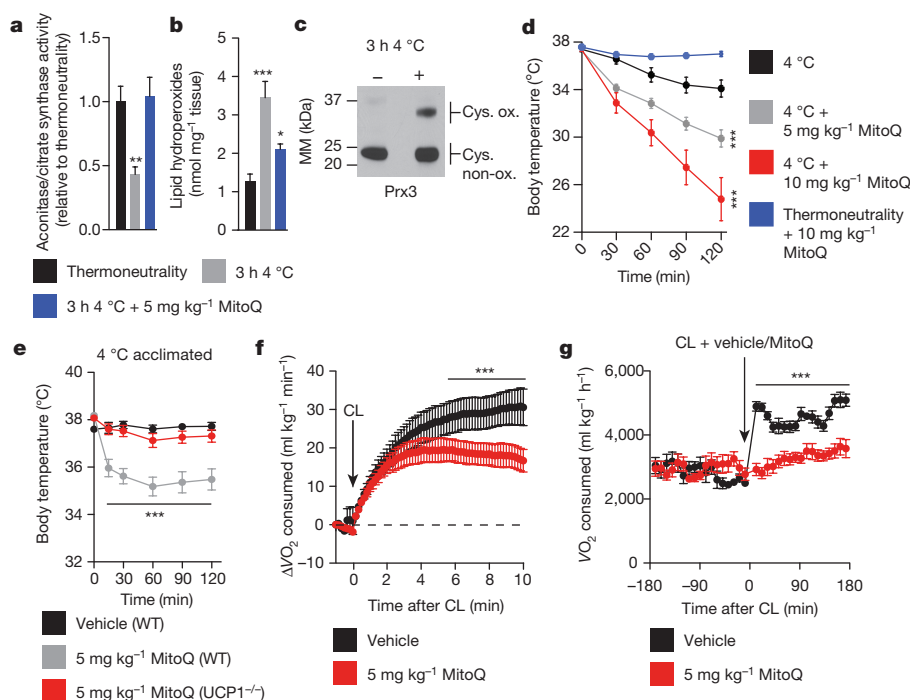


Figure 1 | Increased BAT mitochondrial ROS levels support UCP1-dependent thermogenesis *in vivo*. **a–c**, Effect of acute cold exposure on *in vivo* BAT (**a**) mitochondrial superoxide-dependent inactivation of mitochondrial aconitase ($n = 5$; 5 mg kg^{-1} MitoQ $n = 4$), (**b**) lipid hydroperoxide content ($n = 5$), and (**c**) mitochondrial hydrogen-peroxide-dependent oxidation of Prx3. Oxidized Prx3 was assessed using the redox gel shift method described in Methods. MM, molecular mass. Cys. ox., cysteine oxidation. For uncropped scans see Supplementary Figs 1–3. **d**, Effect of i.p. MitoQ on core body temperature after acute

cold exposure ($n = 10$). **e**, Effect of i.p. MitoQ on core body temperature of WT and UCP1^{-/-} mice after 4 °C acclimation ($n = 10$; UCP1^{-/-} $n = 8$). **f**, Oxygen consumed acutely before and after i.p. CL \pm MitoQ ($n = 5$). **g**, Oxygen consumed 3 h before and after i.p. CL \pm MitoQ ($n = 8$). Data are mean \pm s.e.m. of at least four mouse replicates. * $P < 0.05$, *** $P < 0.001$ (two-tailed Student's *t*-test for pairwise comparisons, one-way/two-way analysis of variance (ANOVA) for multiple comparisons involving one/two independent variables).

We examined this possibility using *N*-acetylcysteine (NAC), a cell-permeable cysteine precursor shown to increase the intracellular pool of reduced thiols¹³. Similar to what was observed with MitoQ, NAC treatment resulted in a significant decrease in core body temperature upon cold exposure (Fig. 2d). This hypothermic activity suggests an effect on BAT function, as shivering was unaffected (Extended Data Fig. 1c–e). Furthermore, acute elevation in oxygen consumption by CL administration was inhibited by NAC pretreatment (Fig. 2e and Extended Data Fig. 2d). Taken together, these findings suggest that increased mitochondrial ROS levels and consequent thiol oxidation play important roles in driving BAT thermogenesis *in vivo*.

The acute hypothermic effects of depleting mitochondrial ROS and altering thiol redox status occurred without any appreciable decrease in thermogenic gene expression in either BAT or inguinal white adipose tissue (Extended Data Fig. 3a, b). In fact, some thermogenic gene markers were further elevated following MitoQ or NAC treatment, possibly representing a compensatory increase in expression due to functional compromise of thermogenic respiration (Extended Data Fig. 3a, b). Furthermore, stimulation of the lipolytic machinery in BAT was comparable upon cold challenge after MitoQ treatment, suggesting this upstream driver of thermogenesis was not compromised (Extended Data Fig. 3c, d). To test whether mitochondrial ROS might act as direct modulators of UCP1-dependent leak respiration, we examined primary brown adipocytes using cellular respirometry^{14,15}. As expected, adrenergic stimulation by noradrenaline triggered a substantial increase in oxygen consumption rate (OCR; Fig. 3a and Extended Data Fig. 3e); while inhibition of ATP synthase with oligomycin allowed for determination of leak respiration^{14,15} (Fig. 3a and Extended Data Fig. 3e). Depletion of mitochondrial ROS with MitoQ and reducing the thiol redox status with NAC both inhibited leak respiration (Fig. 3a, b and Extended Data Fig. 3e, f). Notably, these

inhibitory effects were lost in adipocytes lacking UCP1 (Fig. 3c, d and Extended Data Fig. 3g, h).

Our findings indicated that ROS-supported thermogenesis relied on both UCP1 and cysteine thiol oxidation. So, we considered whether UCP1 cysteines were targets for oxidative modification during thermogenesis. We used a redox gel-shift method¹⁶ to determine whether BAT mitochondrial ROS drive cysteine-centred oxidative modifications on UCP1 *in vivo*. This approach labels reversibly oxidized protein cysteines with a polyethylene glycol maleimide, allowing for their determination by immunoblotting after SDS–polyacrylamide gel electrophoresis¹⁶ (Extended Data Fig. 3i). In BAT from mice housed at thermoneutrality, UCP1 existed predominantly in a cysteine non-oxidized form (Fig. 3e). However, upon cold challenge cysteine oxidation increased on UCP1 in a time-dependent manner (Fig. 3e). Calibration of UCP1 cysteine shifts indicated that a single cysteine site was redox modified during thermogenesis (Extended Data Fig. 4a–c), and this modification did not appear to participate in intermolecular disulfide bridging (Extended Data Fig. 4d). Notably, pharmacological reduction of BAT thiol status with NAC, which results in hypothermia (Fig. 2d), diminished UCP1 cysteine thiol oxidation upon cold exposure (Fig. 3f).

So, ROS-supported thermogenesis in BAT both depended on UCP1 (Fig. 1e) and acted directly on UCP1 cysteine(s) by oxidative modification (Fig. 3e, f). To examine the role of cysteine oxidation on UCP1 function, we first identified the site modified during thermogenesis. We profiled UCP1 cysteines for sulfenic acid modifications *in vivo* by selective dimedone labelling and mass spectrometry¹⁷ (Extended Data Fig. 4e) after acute cold exposure. We identified six of seven UCP1 cysteines, five of which were identified exclusively in the unmodified (*N*-ethyl maleimide (NEM)-labelled) form (Extended Data Fig. 4f and Supplementary Data File 1). However, UCP1 Cys253 was identified in the sulfonylated (dimedone-labelled) form (Fig. 4a, Extended Data Fig. 4f

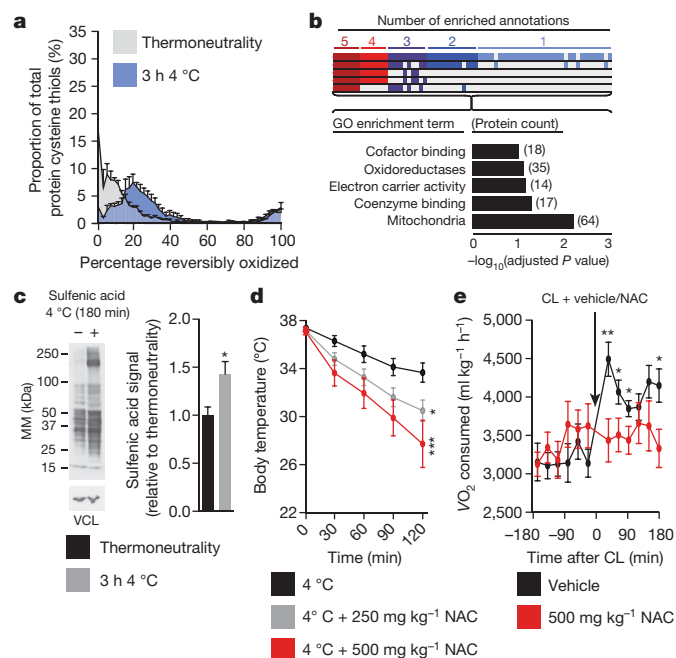


Figure 2 | BAT mitochondrial ROS during thermogenesis drives oxidation of cellular and mitochondrial thiols. **a**, Distribution of percentage reversible oxidation status of BAT protein thiols \pm acute cold exposure. **b**, Pathway analysis of BAT proteins containing cysteine residues sensitive to substantial oxidation ($>10\%$ shift in oxidation status) upon cold exposure. Top: proteins clustered according to shared GO enrichment terms. Bottom: significantly enriched pathways. **c**, Immunodetection of protein sulfenic acid levels in BAT \pm acute cold exposure ($n=4$). **d**, Effect of i.p. NAC on core body temperature after acute cold exposure ($n=8$; 500 mg kg⁻¹ NAC $n=7$). **e**, Oxygen consumed 3 h before and after i.p. CL \pm NAC (control $n=12$; NAC $n=9$). VCL, vinculin. Data are mean \pm s.e.m. of at least four mouse replicates. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (two-tailed Student's *t*-test for pairwise comparisons, one-way/two-way ANOVA for multiple comparisons involving one/two independent variables).

and Supplementary Data File 1). Moreover, sulfenic acid occupancy on Cys253 increased significantly upon acute cold exposure (Fig. 4b). Quantification of dimedone-labelled Cys253 provided an estimate of sulfenylation status during cold exposure (8.4%; Fig. 4b). However, owing to the unstable nature of protein sulfenic acids¹⁸ this must be considered a significant under-representation of *in vivo* levels of this modification.

Since UCP1 Cys253 was identified as a redox-sensitive site for which modification correlated with thermogenesis, we investigated the potential structural and functional importance of this residue. We modelled UCP1 on the structure of the homologous mitochondrial ATP/ADP carrier (AAC)¹⁹ (Fig. 4c and Extended Data Fig. 5a). This structure represents the matrix-closed 'c-state' of the carrier¹⁹, which for UCP1 is maintained by inhibitory binding of cytosolic purine nucleotides^{1,20}.

Examining the environment of the Cys253 thiol indicated that it resides at an interface between two matrix-facing helices. These helices contain residues highly conserved across the carrier family that stabilize the 'c-state' by inter-helical hydrophobic interactions and a hydrogen bond^{21,22} (Fig. 4d and Extended Data Fig. 5b). This region forms the opposing face of the cytosolic cavity where inhibitory nucleotide binding is thought to occur^{23,24}, and is proximal to a putative site of cardiolipin binding on UCP1²¹ (Fig. 4d). Notably, modelling a sulfenic acid modification on the Cys253 thiol suggested that these 'c-state' interactions could be destabilized owing to local electrostatic effects (Extended Data Fig. 5b). Thus, on the basis of this structural model, Cys253 oxidation status could play a role in determining stability of the matrix closed 'c-state' of UCP1, and sensitivity to β -adrenergic activation and uncoupling.

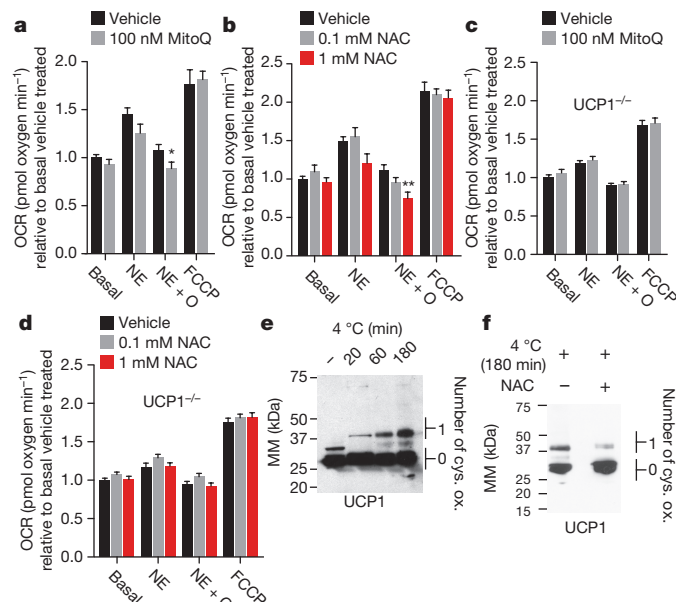


Figure 3 | BAT mitochondrial ROS oxidatively modify a cysteine residue on UCP1 and support UCP1-dependent leak respiration.

a, **b**, OCR of brown adipocytes \pm noradrenaline (NE) stimulation + oligomycin to determine leak respiration \pm (a) MitoQ ($n=10$) or (b) NAC ($n=8$; 0.1 mM NAC $n=7$). **c**, **d**, OCR of primary UCP1^{-/-} adipocytes \pm noradrenaline stimulation + oligomycin \pm (c) MitoQ ($n=10$) or (d) NAC ($n=10$). **e**, **f**, Cys-redox immunoblot of BAT UCP1 (e) after a time course of acute cold exposure, and (f) after acute cold exposure \pm 500 mg kg⁻¹ NAC. FCCP, carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone. Data are mean \pm s.e.m. of at least seven cellular replicates. * $P < 0.05$, ** $P < 0.01$ (two-tailed Student's *t*-test for pairwise comparisons, one-way ANOVA for multiple comparisons).

In the cellular milieu, UCP1 is inhibited by endogenous purine nucleotides, and this inhibition can be overcome by β -adrenergic stimulation of lipolysis and activation of UCP1-dependent respiration^{1,15}. A prediction of the above model is that altering Cys253 status should specifically modulate sensitivity to adrenergic activation of the purine nucleotide inhibited 'c-state' of UCP1^{1,20}. To test this, in UCP1^{-/-} brown adipocytes we separately transduced WT UCP1 and seven mutants, each having one of the seven cysteine residues contained in the protein replaced with an alanine. We titrated UCP1 transduction to recapitulate RNA and protein abundance observed in WT brown adipocytes (Extended Data Figure 6a–d). In all cases, basal and maximal respiration was indistinguishable between UCP1^{-/-} brown adipocytes, UCP1^{-/-} adipocytes transduced with WT UCP1, and UCP1^{-/-} adipocytes transduced with each of the seven UCP1 Cys-mutants (Extended Data Fig. 6e, f). This agreed with previous studies indicating that UCP1 cysteines are not critical for protein stability and mitochondrial viability²⁵.

The UCP1 Cys-mutant cells were subjected to noradrenaline and oligomycin to examine the effect of adrenergic activation on leak respiration. Noradrenaline stimulated rapid oxidation of protein cysteine residues (Extended Data Fig. 6g) that was concomitant with increases in OCR (Extended Data Fig. 6h). As expected, adipocytes lacking UCP1 exhibited a substantial reduction in respiration compared with WT UCP1 cells after 100 nM noradrenaline and oligomycin treatment, with $\sim 35\%$ of total leak OCR determined to be UCP1-dependent (Fig. 4e). Although replacing most UCP1 cysteines with alanine sites had no effect (Extended Data Fig. 7a–e), individual modification of Cys253 substantially inhibited UCP1-dependent leak respiration (Fig. 4f). Interestingly, individual modification of Cys224 also inhibited UCP1 leak respiration (Extended Data Fig. 7f). We additionally generated a C224A/C253A double mutant (Extended Data Fig. 6c, d), which exhibited compromise of UCP1-dependent

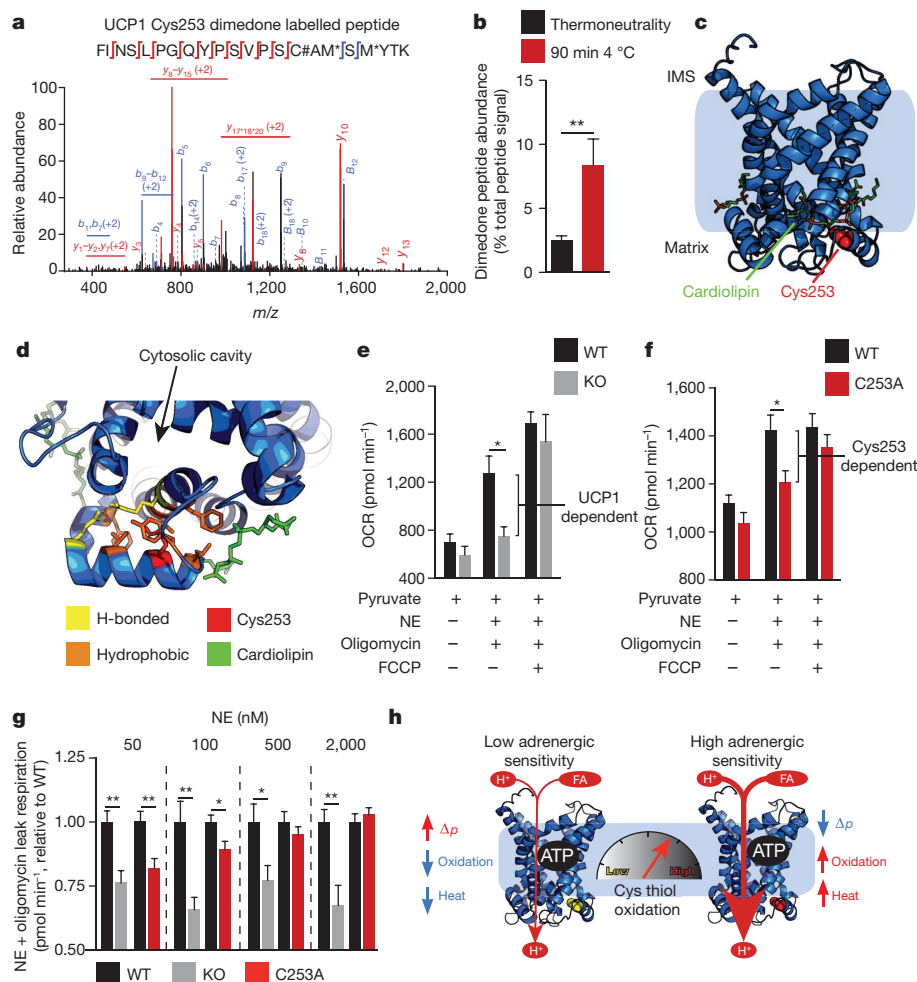


Figure 4 | UCP1 Cys253 is sulfenylated during thermogenesis and sensitizes UCP1 to adrenergic activation. **a**, MS² spectrum of dimedone-labelled UCP1 Cys253 peptide indicating sulfenylation of this site during thermogenesis. Fragment ions that span the dimedone-alkylated cysteine are highlighted in the peptide sequence. C#, dimedone-labelled cysteine; M*, oxidized methionine. **b**, Quantification of dimedone-labelled UCP1 Cys253 relative to the NEM-alkylated form (*n* = 5). **c**, **d**, Structure of (c) human UCP1 modelled on the AAC crystal structure and (d) Cys253 in a hydrophobic pocket between two matrix facing helices. IMS, intermembrane space. **e**, **f**, Basal, maximal, and UCP1-dependent OCR of UCP1^{-/-} brown adipocytes ± transduction with (e) WT UCP1 (WT *n* = 7; UCP1^{-/-} *n* = 6) or (f) UCP1 C253A (WT *n* = 17; C253A *n* = 19).

respiration comparable to the individual mutations (Extended Data Fig. 7g).

Replacement of the redox-sensitive Cys253 inhibited adrenergic activation of UCP1 by noradrenaline, so we next tested whether this site played an obligatory or sensitizing role in this context. To do so, we titrated noradrenaline (50–2,000 nM) to trigger variable levels of adrenergic stimulation in brown adipocytes (Extended Data Fig. 8a). All concentrations of noradrenaline were sufficient to activate WT UCP1-dependent leak respiration, which was consistently ~25–35% of total leak respiration (Fig. 4g and Extended Data Fig. 8b). UCP1-dependent respiration in C253A cells was substantially inhibited after stimulation with 50 nM noradrenaline; however, this inhibition was overcome dose-dependently by increasing noradrenaline concentrations (Fig. 4g). Thus, increased adrenergic stimulus could overcome inhibition by Cys253 replacement, suggesting that Cys253 plays a sensitizing (as opposed to an obligatory) role in adrenergically stimulated uncoupled respiration by UCP1. In contrast, exposure to variable adrenergic stimulus did not substantially modify the degree of inhibition of UCP1-dependent leak respiration in UCP1 C224A cells (Extended Data

KO, knockout. **g**, UCP1-dependent leak respiration after stimulation by various concentrations of noradrenaline (50 nM noradrenaline *n* = 9; 100 nM noradrenaline *n* = 7; 500 nM noradrenaline *n* = 8; 2,000 nM noradrenaline *n* = 6). Comparison of UCP1 WT and C253A indicates that degree of UCP1 inhibition by C253A is inversely correlated with noradrenaline concentration (*n* = 19; 100 nM WT noradrenaline; *n* = 17,500 nM noradrenaline *n* = 18; 2,000 nM noradrenaline *n* = 10). **j**, A model of sensitization of UCP1-mediated uncoupling by mitochondrial ROS. Data are mean ± s.e.m. of at least five mouse replicates or cell replicates for respirometry experiments. **P* < 0.05, ***P* < 0.01 (two-tailed Student's *t*-test for pairwise comparisons).

Fig. 8c). So, as opposed to a specific effect through altering sensitivity to adrenergic activation as in the case of Cys253, Cys224 replacement may alter UCP1 structure/function more generally.

Moreover, permeabilizing adipocytes to dilute cytosolic purine nucleotides resulted in uncoupled respiration that was indistinguishable between WT and C253A cells (Extended Data Fig. 8d), and could be abrogated by sequestering of endogenous free fatty acids with BSA (Extended Data Fig. 8d). In this nucleotide-diluted state, WT and C253A UCP1 cells exhibited a comparably low membrane potential ($\Delta\psi$) as assessed by tetramethylrhodamine methyl ester (TMRM) fluorescence (Extended Data Figure 8e). As expected, titration of GDP resulted in an increase in $\Delta\psi$ in UCP1 WT cells owing to GDP-dependent inhibition of UCP1^{1,20} (Extended Data Fig. 8f). Notably, the C253A and C224A/C253A mutants exhibited increased sensitivity to GDP-dependent inhibition (Extended Data Fig. 8f).

Taken together, these data indicate that acute regulation of thermogenesis *in vivo* can be supported by elevation of mitochondrial ROS levels. The thermogenic actions of mitochondrial ROS involve consequent oxidation of BAT cellular thiols, and UCP1—the major

thermogenic effector—is a target of thiol oxidation. We identify UCP1 Cys253 as a sulfenylation site that is increasingly modified during thermogenesis, and provide evidence that this residue can modulate sensitivity to activation of UCP1-dependent thermogenic respiration.

While UCP1 is identified here as a target of redox modification during thermogenesis, it is likely that other functional target proteins exist, which may act together to support thermogenic respiration. In addition to establishing a role for thiol redox signalling in thermogenesis *in vivo*, these findings provide insight into the longstanding investigations on the relationship between mitochondrial ROS and UCP1 (refs 1, 2, 26, 27). Interestingly, studies in this area have relied largely on the use of isolated mitochondria and reconstituted systems. Our data suggest that redox regulation of UCP1-dependent respiration is particularly relevant in adipocytes and *in vivo* where both redox homeostasis and free fatty-acid concentrations are subject to strict endogenous regulation; parameters that are necessarily divergent from *in vitro* conditions. Notably, a role for thiol redox status and UCP1-dependent uncoupling has been suggested in isolated BAT mitochondria²⁸, while the importance of superoxide in particular for modulating UCP1 function is the subject of debate^{1,2,26,27}. Previous investigation using a manganese superoxide dismutase transgenic mouse model suggested that matrix superoxide did not modify UCP1 function²⁷. This previous work suggests that the role of thiol redox status in UCP1-dependent respiration detailed here is not directly mediated by matrix superoxide, but reliant on other ROS and related molecules the levels of which are not compromised by elevated manganese superoxide dismutase expression¹⁰.

Our data are consistent with a model whereby redox modification of Cys253 sensitizes purine nucleotide inhibited UCP1 to uncoupling due to acute adrenergic stimulus (Fig. 4j), which is compatible with fatty-acid-mediated activation and shuttling²⁰. This model reconciles the apparent sensitizing role of both redox status and UCP1 Cys253 with the sufficient role of fatty acids for UCP1-dependent uncoupling¹, as well as previous observations that cysteines are not essential for UCP1 activity in reconstituted systems²⁵. However, previous study of UCP1 cysteines used yeast expression systems that are subject to protein misfolding and artefactual uncoupling²⁹. Although Cys224 was not identified here as a target for reversible redox modification it appears to affect UCP1-dependent respiration, albeit distinctly from the sensitizing role of Cys253. It could be that this site is modified by other ROS-dependent species that are known to interact with UCP1 cysteines³⁰. Additionally, these sites may also be important for UCP1 structure/function independent of redox-modification. Finally, it is worth noting that UCP1 Cys253 is conserved to a high degree across several mitochondrial carriers, including UCP2, UCP3, and AAC. This suggests that redox modification of this site may be a pervasive signalling modality for mitochondrial carriers in physiology and pathophysiology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 July 2015; accepted 4 February 2016.

Published online 30 March 2016.

- Nicholls, D. G. The physiological regulation of uncoupling proteins. *Biochim. Biophys. Acta* **1757**, 459–466 (2006).
- Divakaruni, A. S. & Brand, M. D. The regulation and physiology of mitochondrial proton leak. *Physiology* **26**, 192–205 (2011).
- Yoneshiro, T. et al. Recruited brown adipose tissue as an antiobesity agent in humans. *J. Clin. Invest.* **123**, 3404–3408 (2013).
- Wu, J. et al. Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. *Cell* **150**, 366–376 (2012).
- Cox, A. G., Peskin, A. V., Paton, L. N., Winterbourn, C. C. & Hampton, M. B. Redox potential and peroxide reactivity of human peroxiredoxin 3. *Biochemistry* **48**, 6495–6501 (2009).
- Rodriguez-Cuenca, S. et al. Consequences of long-term oral administration of the mitochondria-targeted antioxidant MitoQ to wild-type mice. *Free Radic. Biol. Med.* **48**, 161–172 (2010).
- Cannon, B. & Nedergaard, J. Nonshivering thermogenesis and its adequate measurement in metabolic studies. *J. Exp. Biol.* **214**, 242–253 (2011).

- Ukropec, J., Anunciado, R. P., Ravussin, Y., Hulver, M. W. & Kozak, L. P. UCP1-independent thermogenesis in white adipose tissue of cold-acclimated *Ucp1*^{−/−} mice. *J. Biol. Chem.* **281**, 31894–31908 (2006).
- Himms-Hagen, J. et al. Effect of CL-316,243, a thermogenic beta 3-agonist, on energy balance and brown and white adipose tissues in rats. *Am. J. Physiol.* **266**, R1371–R1382 (1994).
- Collins, Y. et al. Mitochondrial redox signalling at a glance. *J. Cell Sci.* **125**, 801–806 (2012).
- Seo, Y. H. & Carroll, K. S. Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies. *Proc. Natl Acad. Sci. USA* **106**, 16163–16168 (2009).
- Finkel, T. From sulfenylation to sulfhydration: what a thiolate needs to tolerate. *Sci. Signal.* **5**, pe10 (2012).
- Atkuri, K. R., Mantovani, J. J., Herzenberg, L. A. & Herzenberg, L. A. N-Acetylcysteine—a safe antidote for cysteine/glutathione deficiency. *Curr. Opin. Pharmacol.* **7**, 355–359 (2007).
- Wikstrom, J. D. et al. Hormone-induced mitochondrial fission is utilized by brown adipocytes as an amplification pathway for energy expenditure. *EMBO J.* **33**, 418–436 (2014).
- Li, Y., Fromme, T., Schweizer, S., Schöttl, T. & Klingenspor, M. Taking control over intracellular fatty acid levels is essential for the analysis of thermogenic function in cultured primary brown and brite/beige adipocytes. *EMBO Rep.* **15**, 1069–1076 (2014).
- Requejo, R. et al. Measuring mitochondrial protein thiol redox state. *Methods Enzymol.* **474**, 123–147 (2010).
- Martínez-Acedo, P., Gupta, V. & Carroll, K. S. Proteomic analysis of peptides tagged with dimedone and related probes. *J. Mass Spectrom.* **49**, 257–265 (2014).
- Nelson, K. J. et al. Use of dimedone-based chemical probes for sulfenic acid detection methods to visualize and identify labeled proteins. *Methods Enzymol.* **473**, 95–115 (2010).
- Pebay-Peyroula, E. et al. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* **426**, 39–44 (2003).
- Fedorenko, A., Lishko, P. V. & Kirichok, Y. Mechanism of fatty-acid-dependent UCP1 uncoupling in brown fat mitochondria. *Cell* **151**, 400–413 (2012).
- Lee, Y., Willers, C., Kunji, E. R. & Crichton, P. G. Uncoupling protein 1 binds one nucleotide per monomer and is stabilized by tightly bound cardiolipin. *Proc. Natl Acad. Sci. USA* **112**, 6973–6978 (2015).
- Ruprecht, J. J. et al. Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. *Proc. Natl Acad. Sci. USA* **111**, E426–E434 (2014).
- Modrianský, M., Murdza-Inglis, D. L., Patel, H. V., Freeman, K. B. & Garlid, K. D. Identification by site-directed mutagenesis of three arginines in uncoupling protein that are essential for nucleotide binding and inhibition. *J. Biol. Chem.* **272**, 24759–24762 (1997).
- Divakaruni, A. S., Humphrey, D. M. & Brand, M. D. Fatty acids change the conformation of uncoupling protein 1 (UCP1). *J. Biol. Chem.* **287**, 36845–36853 (2012).
- Arechaga, I. et al. Cysteine residues are not essential for uncoupling protein function. *Biochem. J.* **296**, 693–700 (1993).
- Echtay, K. S. et al. Superoxide activates mitochondrial uncoupling proteins. *Nature* **415**, 96–99 (2002).
- Silva, J. P. et al. SOD2 overexpression: enhanced mitochondrial tolerance but absence of effect on UCP activity. *EMBO J.* **24**, 4061–4070 (2005).
- Mailloux, R. J., Adjeitey, C. N., Xuan, J. Y. & Harper, M. E. Crucial yet divergent roles of mitochondrial redox state in skeletal muscle vs. brown adipose tissue energetics. *FASEB J.* **26**, 363–375 (2012).
- Stuart, J. A., Harper, J. A., Brindle, K. M., Jekabsons, M. B. & Brand, M. D. A mitochondrial uncoupling artifact can be caused by expression of uncoupling protein 1 in yeast. *Biochem. J.* **356**, 779–789 (2001).
- Malingieux, E. A. et al. Fatty acids are key in 4-hydroxy-2-nonenal-mediated activation of uncoupling proteins 1 and 2. *PLoS ONE* **8**, e77786 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements Supported by the JPB Foundation and National Institutes of Health (DK31405) to B.M.S. and by grants from Human Frontiers Science Program (to E.T.C.) and the Canadian Institutes for Health Research (to L.K.). We acknowledge M. Murphy, Y. Kirichok, and A. Bertholet for many discussions. We also thank M. Murphy for providing MitoQ.

Author Contributions E.T.C. designed research, performed biochemical, cellular, and *in vivo* experiments, analysed data, and co-wrote the paper. L.K. designed and performed cellular and mutagenesis experiments. M.P.J. and K.A.P. performed and analysed data from mass spectrometric experiments. G.Z.L. performed cellular experiments. C.B.C. and S.P.G. oversaw mass spectrometric experiments. A.J.R. designed and performed structural modelling. E.T.C. and B.M.S. directed the research and co-wrote the paper, with assistance from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.M.S. (bruce_spiegelman@dfci.harvard.edu).

METHODS

Animal procedures and ethics statement. Animal experiments were performed according to procedures approved by the Institutional Animal Care and Use Committee of the Beth Israel Deaconess Medical Center. Unless otherwise stated, mice used were male C57BL/6J (8–12 weeks of age; Jackson Laboratories), and housed in a temperature-controlled (20–22°C) room on a 12 h light/dark cycle.

***In vivo* compound administration.** All compounds administered to mice *in vivo* were injected at the stated dose i.p. 10 min before subsequent interventions unless otherwise stated.

Body temperature and cold exposure. Body temperature and cold exposure experiments were assessed using a mouse rectal probe (World Precision Instruments). When studying acute activation of thermogenesis, mice were housed from birth at 20–22°C to allow for recruitment of thermogenic adipose tissue⁷. Before individual housing at 4°C, mice were placed at thermoneutrality (30°C) for 3 days which allows both for maintenance BAT UCP1 protein content³¹ and for measurement of acute induction of BAT thermogenesis upon cold exposure. Upon exposure to 4°C, temperature was measured every 30 min. When studying body temperature after 4°C acclimation, WT and Ucp1^{−/−} mice (equal numbers of male and female mice in each group) were acclimated using established protocols: mice were individually housed for 1 week at 15°C, 1 week at 10°C, and 24 h at 4°C before the experiment.

EMG determination of muscle shivering. Mice were individually restrained to limit non-shivering muscle activity and two EMG needle electrodes were inserted subcutaneously above the nuchal muscles in the back of the neck. EMG leads were connected to a computerized data acquisition system via a communicator. EMG was recorded at thermoneutrality to determine non-shivering basal nuchal muscle activity, before placement of mice at 4°C. EMG data were collected and burst activity was determined as described previously³². Briefly, EMG data were collected from the implanted electrodes at a sampling rate of 2 kHz using LabChart 8 Pro Software (ADInstruments). The raw signal was converted to root mean square activity. Root mean square activity was analysed for shivering bursts in 10 s windows.

Metabolic phenotyping. Whole-body energy metabolism was evaluated using a Comprehensive Lab Animal Monitoring System (CLAMS, Columbia Instruments). For 6 h measurements, mice were acclimated in the metabolic chambers for 48 h before experiments to minimize stress from the housing change. CO₂ and O₂ levels were collected every 12 or 32 min for each mouse over the period of the experiment. For acute measurements, CO₂ and O₂ levels were collected every 10 s. CL 316,243 (Sigma-Aldrich; 1 mg kg^{−1}) was injected i.p. into mice at the indicated times.

Assessment of mitochondrial aconitase inactivation in mouse BAT. Aconitase activity was measured as described previously³³. In brief, after the relevant *in vivo* intervention mouse BAT was rapidly excised and homogenized in mitochondrial isolation buffer (250 mM sucrose, 2 mM EDTA, 10 mM sodium citrate, 0.6 mM MnCl₂, 100 mM Tris-HCl, pH 7.4) followed by mitochondrial isolation by differential centrifugation. Samples (1–2 mg mitochondrial protein) were added to a 96-well plate and 190 µl assay buffer (50 mM Tris-HCl (pH 7.4), 0.6 mM MnCl₂, 5 mM sodium citrate, 0.2 mM NADP⁺, 0.1% (v/v) Triton X-100, 0.4 U ml^{−1} ICDH). Absorbance was measured at 340 nm for 7 min at 37°C. To control for mitochondrial content aconitase activity was normalized to citrate synthase activity³⁴ and expressed the result as a percentage of control levels.

Assessment of lipid hydroperoxides in mouse BAT. Lipid hydroperoxide content in mouse BAT was estimated by rapid snap freezing of BAT tissue followed by lipid extraction and assessment using a modified ferric thiocyanate assay (Cayman Chemical Lipid Hydroperoxide Assay Kit) according to the manufacturer's instructions.

Assessment of cysteine redox status by gel shift. Cysteine redox status of Prx3 and UCP1 was measured as described previously^{16,35}. After the relevant *in vivo* intervention, mouse BAT was rapidly excised and homogenized in 100 mM NEM, 1 mM EGTA, 50 mM Tris-HCl, pH 7.4. Samples were incubated at 37°C for 5 min before the addition of SDS (2% final) and further incubation at 37°C for 10 min. Incubations at 37°C proceeded in a thermomixer at 1,300 r.p.m. Samples were then precipitated in five volumes of ice-cold acetone to remove excess NEM before resuspension in 1 mM EGTA, 2% SDS, 10 mM TCEP, 50 mM Tris-HCl, pH 7.4 containing a polyethylene glycol polymer conjugated to maleimide (50 mM PEG-Mal). Resuspended samples were incubated for 30 min at 37°C before a second acetone precipitation to remove excess PEG-Mal before sample resuspension and immunoblot detection by standard methods described below. For UCP1 experiments, to ensure gel shift signals were specific to reversible cysteine oxidation, oxidized samples were separately treated with TCEP before differential labelling as described above. Calibrating the number of UCP1 cysteines oxidized was achieved by treating TCEP-reduced samples with increasing proportions of Peg-Mal:NEM to generate a

cysteine-dependent ladder³⁵. In addition, to ensure higher molecular mass signals were specific to UCP1, UCP1 antibody specificity was tested in BAT. It should be noted that while the UCP1 antibody used here is highly specific for UCP1 in BAT (Extended Data Fig. 4c), the same antibody applied to cultured brown adipocyte samples can generate non-specific signals at molecular mass >35 kDa. So, the UCP1 gel shift assay as described here is only compatible with *in vivo* tissue experiments.

Assessment of BAT-reduced and -oxidized glutathione content by mass spectrometry. Reduced and oxidized glutathione were profiled in negative ionization mode by liquid chromatography tandem mass spectrometry (LC-MS) methods as described previously³⁶. Data were acquired using an ACQUITY UPLC (Waters) coupled to a 5500 QTRAP triple quadrupole mass spectrometer (AB SCIEX). Tissue homogenates (30 µl) were extracted using 120 µl of 80% methanol containing 0.05 ng µl^{−1} inosine-¹⁵N₄, 0.05 ng µl^{−1} thymine-d₄, and 0.1 ng µl^{−1} glycocholate-d₄ as internal standards (Cambridge Isotope Laboratories). The samples were centrifuged (10 min, 9,000g, 4°C) and the supernatants (10 µl) were injected directly onto a 150 mm × 2.0 mm Luna NH₂ column (Phenomenex). The column was eluted at a flow rate of 400 µl min^{−1} with initial conditions of 10% mobile phase A (20 mM ammonium acetate and 20 mM ammonium hydroxide (Sigma-Aldrich) in water (VWR)) and 90% mobile phase B (10 mM ammonium hydroxide in 75:25 v/v acetonitrile/methanol (VWR)) followed by a 10 min linear gradient to 100% mobile phase A. The ion spray voltage was −4.5 kV and the source temperature was 500°C. Raw data were processed using MultiQuant 2.1 software (AB SCIEX) for automated peak integration. LC-MS data were processed and visually inspected using TraceFinder 3.1 software (Thermo Fisher Scientific).

Assessment of BAT protein thiol redox status using iodoTMT. After the relevant *in vivo* intervention, mouse BAT was rapidly excised and homogenized in 20% (w/v) TCA to stabilize thiols. The homogenate was incubated on ice for 30 min and then pelleted for 30 min at 16,000g at 4°C. The pellet was washed with 10% and 5% (w/v) TCA and then resuspended in 80 µl denaturing alkylating buffer (DAB; 6 M urea, 2% (w/v) SDS, 200 mM Tris-HCl, 10 mM EDTA, 100 µM DTPA, 10 µM neocuproine). The contents of one vial of iodoTMT₆ reagent (Thermo Scientific) was added to each of three biological replicate samples to label reduced cysteine residues at 37°C and 1,300 r.p.m. for 1 h. Sample protein was precipitated with five volumes of ice-cold acetone, incubated at −20°C for 2 h, and pelleted at 4°C and 16,000g for 30 min. The amount of protein to be processed was optimized to ensure saturation of thiol labelling by the iodoTMT reagent as per the manufacturer's instructions. The pellet was washed twice with ice-cold acetone and then re-solubilized in 80 µl DAB containing 1 mM tris(2-carboxyethyl)phosphine (TCEP), reducing previously reversibly oxidised cysteine residues in the presence of a second, distinct iodoTMT₆ reagent. Proteins were incubated at 37°C and 1,400 r.p.m. for 1 h, precipitated and resuspended for protease digestion. After digestion, iodoTMT-labelled cysteine-containing peptides were enriched using the anti-TMT resin as per the manufacturer's instructions.

Profiling of redox-sensitive protein targets. Proteins with cysteine thiols exhibiting differential redox status (defined as >10% shift in cysteine oxidation status upon cold exposure) were assessed for Gene Ontology (GO) term enrichment³⁷. The total identified population of cysteine thiol containing proteins was used as the reference background. Enriched GO terms were filtered after benjamini-hochberg correction at an adjusted *P* value <0.1. All data analysis used R (R Core Team, Vienna, Austria, <http://www.R-project.org>).

Assessment of protein thiol sulfenic acids. Tissue or cellular samples were prepared adapting a protocol used previously to stabilize endogenous protein sulfenic acids³⁸. Briefly, samples were homogenized in 50 mM Tris base, containing 100 mM NaCl, 100 µM DTPA, 0.1% SDS, 0.5% sodium deoxycholate, 0.5% Triton-X 100, 5 mM dimedone. To minimize lysis-dependent oxidation, buffers were bubbled with argon before use. Samples were incubated for 15 min at room temperature, at which point SDS was added to a final concentration of 1% and samples were incubated for a further 15 min. After dimedone treatment, 10 mM TCEP and 50 mM NEM were added and samples were incubated for a further 15 min at 37°C to reduce and alkylate all non-sulfenic acid protein cysteine residues. Protein sulfenic acids were then assessed by immunoblotting against dimedone (1:1,000 antibody dilution).

Targeted assessment of UCP1 cysteine sulfonylation status using dimedone-MS. After dimedone and NEM labelling of samples as described above, samples were resolved by SDS-PAGE and bands in the UCP1 containing region of the gel (30–35 kDa) were excised, destained with acetonitrile and subjected to dehydration by a speed vacuum concentrator. Gel bands were rehydrated with digestion buffer (75 µl of 50 mM HEPES and 500 ng of trypsin (Promega) and subjected to 12 h of digestion at 37°C. Peptides were extracted and labelled with TMT 10 reagents (Thermo Fisher) as previously described³⁹.

Protein digestion. Protein pellets were dried and resuspended in 8 M urea containing 50 mM HEPES (pH 8.5). Protein concentrations were measured by BCA assay (Thermo Scientific) before protease digestion. Protein lysates were diluted to 4 M urea and digested with LysC (Wako, Japan) in a 1/100 enzyme/protein ratio overnight. Protein extracts were diluted further to a 1.0 M urea concentration, and trypsin (Promega) was added to a final 1/200 enzyme/protein ratio for 6 h at 37°C. Digests were acidified with 20 μ l of 20% formic acid (FA) to a pH \sim 2, and subjected to C18 solid-phase extraction (Sep-Pak, Waters).

LC-MS/MS parameters. All spectra were acquired using an Orbitrap Fusion mass spectrometer (Thermo Fisher) in line with an Easy-nLC 1000 (Thermo Fisher Scientific) ultra-high pressure liquid chromatography pump. Peptides were separated onto a 100 μ m inner diameter column containing 1 cm of Magic C4 resin (5 μ m, 100 Å, Michrom Bioresources) followed by 30 cm of Sepax Technologies GP-C18 resin (1.8 μ m, 120 Å) with a gradient consisting of 9–30% (ACN, 0.125% FA) over 180 min at \sim 250 nL min⁻¹. For all LC-MS/MS experiments, the mass spectrometer was operated in the data-dependent mode. We collected MS¹ spectra at a resolution of 120,000, with an AGC target of 150,000 and a maximum injection time of 100 ms. The ten most intense ions were selected for MS² (excluding 1 Z-ions). MS¹ precursor ions were excluded using a dynamic window (75 s \pm 10 ppm). The MS² precursors were isolated with a quadrupole mass filter set to a width of 0.5 Th. For the MS³ based TMT quantitation, MS² spectra were collected at an AGC of 4,000, maximum injection time of 150 ms, and CID collision energy of 35%. MS³ spectra were acquired with the same Orbitrap parameters as the MS² method except HCD collision energy was increased to 55%. Synchronous-precursor-selection was enabled to include up to six MS² fragment ions for the MS³ spectrum.

Data processing and MS² spectra assignment. A compilation of in-house software was used to convert raw files to mzXML format, as well as to adjust monoisotopic *m/z* measurements and erroneous peptide charge state assignments. Assignment of MS² spectra was performed using the SEQUEST algorithm⁴⁰. All experiments used the Mouse UniProt database (downloaded 10 April 2014) where reversed protein sequences and known contaminants such as human keratins were appended. SEQUEST searches were performed using a 20 ppm precursor ion tolerance, while requiring each peptide's amino/carboxy (N/C) terminus to have trypsin protease specificity and allowing up to two missed cleavages. IodoTMT tags on cysteine residues residues (+329.226595 Da) was set as static modifications, while methionine oxidation (+15.99492 Da) was set as variable modifications. For targeted assessment of UCP1 cysteine sulfonylation, TMT tags on lysine residues and peptide N termini (+229.16293 Da), NEM on cysteine residues (+125.047679 Da) were set as static modifications and oxidation of methionine residues (+15.99492 Da) and dimedone on cysteine residues (+13.020401 Da versus NEM) as variable modifications. Determination of sulfonylation status of the Cys253 peptide was determined by comparing TMT reporter ion abundance of the dimedone-alkylated and NEM-alkylated peptides as a proportion of total precursor ion intensity. An MS² spectra assignment false discovery rate of less than 1% was achieved by applying the target-decoy database search strategy⁴¹. Protein filtering was performed using an in-house linear discrimination analysis algorithm to create one combined filter parameter from the following peptide ion and MS2 spectra metrics: XCorr, Δ Cn score, peptide ion mass accuracy, peptide length and missed-cleavages⁴². Linear discrimination scores were used to assign probabilities to each MS² spectrum for being assigned correctly, and these probabilities were further used to filter the data set to a 1% protein-level false discovery rate.

Determination of iodoTMT reporter ion intensities and quantitative data analysis. For quantification, a 0.03 *m/z* window centred on the theoretical *Th* value of each reporter ion was used for the nearest signal intensity. Reporter ion intensities were adjusted to correct for the isotopic impurities from the different TMT reagents (manufacturer specifications). The signal to noise values for all peptides were summed within each TMT channel. For each peptide, a total minimum sum signal to noise value of 200 and an isolation purity greater than 70% was required⁴³. Percentage cysteine oxidation status of protein thiols was calculated as the percentage of the cysteine containing peptide (total or mitochondrial) labelled with iodoTMT (129, 130, 131) for each condition over the sum of the reduced peptide labelled with iodoTMT (126, 127, 128) plus reversibly oxidized labelled peptide (129, 130, 131): (oxidized peptide 129, 130, 131)/(reduced peptide 126, 127, 128 + oxidized peptide 129, 130, 131) \times 100.

Primary brown adipocyte preparation and differentiation. Interscapular brown adipose stromal vascular fraction was obtained from 2- to 6-day-old pups as described previously⁴⁴. Interscapular brown adipose was dissected, washed in PBS, minced, and digested for 45 min at 37°C in PBS containing 1.5 mg ml⁻¹ collagenase B, 123 mM NaCl, 5 mM KCl, 1.3 mM CaCl₂, 5 mM glucose, 100 mM HEPES, and 4% essentially fatty-acid-free BSA. Tissue suspension was filtered through a 40 μ m cell strainer and centrifuged at 600g for 5 min to pellet the SVF. The cell

pellet was resuspended in adipocyte culture medium and plated. Primary brown pre-adipocytes were counted and plated in the evening, 12 h before differentiation at 15,000 cells per well of a Seahorse plate. Pre-adipocyte plating was scaled according to surface area. The following morning, brown pre-adipocytes were induced to differentiate for 2 days with an adipogenic cocktail (1 μ M rosiglitazone, 0.5 mM IBMX, 5 μ M dexamethasone, 0.114 μ g ml⁻¹ insulin, 1 nM T3, and 125 μ M Indomethacin) in adipocyte culture medium. Two days after induction, cells were re-fed every 48 h with adipocyte culture medium containing 1 μ M rosiglitazone and 0.5 μ g ml⁻¹ insulin. Cells were fully differentiated by day 5 after induction.

Primary brown adipocyte cellular respirometry. Cellular OCR of primary brown adipocytes was determined using a Seahorse XF24 Extracellular Flux Analyzer. Adipocytes were plated and differentiated in XF24 V7 cell culture microplates. Before analysis adipocyte culture medium was changed to DMEM respiration medium lacking NaHCO₃ (Sigma), and including 1.85 g l⁻¹ NaCl, 3 mg l⁻¹ phenol red, 2% fatty-acid-free BSA, 1 mM sodium pyruvate, pH 7.4. Basal respiration was determined to be the OCR in the presence of substrate alone. ATP-synthase-independent respiration was determined after addition of 2.5 μ M oligomycin. Unless otherwise stated, leak respiration was determined after addition of 2.5 μ M oligomycin and 100 nM noradrenaline. Maximal respiration was determined after addition of 2 μ M FCCP. To determine OCR after plasma membrane permeabilization, cells were treated with 50 μ g ml⁻¹ saponin, and sequestration of free fatty acids after permeabilization was achieved through addition of 2% fatty-acid-free BSA.

Cloning, construct design, and site-directed mutagenesis. RNA from murine BAT was reverse-transcribed and used as template for PCR of Ucp1. Sequences for Ucp1 amplification were as follows: sense, CAC CAT GGT GAA CCC GAC AAC TTC C; antisense, TTA TGT GGT ACA ATC CAC TG. PCR fragments were gel-purified and cloned into the pENTR/D-TOPO entry vector according to the manufacturer's instructions (Invitrogen; K2400). Cloned Ucp1 was shuttled into the pAd/CMV/V5-DEST Gateway vector, and confirmed by sequencing. Cysteine mutants were generated using the Quik-Change site-directed mutagenesis kit (Stratagene). Primers for generating mutants were as follows: *Ucp1 C24A* forward 5'-AGCCGGAGTTTCAGCTGCCCTGGCAGATATCATC-3', reverse 5'-GATGATATCTGCCAGGGCAGCTGAAACTCCGGCT-3'; *Ucp1 C188A* forward 5'-TGAGAAATGTCATCATCAATGCTACAGAGCTGGTAACATATG-3', reverse 5'-CATATGTTACCAGCTCTGTAGATTGATGATGACATTTCTCA-3'; *UCP1 C213A* forward 5'-TGGCAGATGACGTCGCCGCCCATTTTACTGTCAGCTC-3', reverse 5'-GAGCTGACAGTAAATGGCGGGGACGTCATCTGCCA-3'; *Ucp1 C224A* forward 5'-TCTTGTGTCGGGGTTTGCCACCACACTCCTGGCC-3', reverse 5'-GGCCAGGAGTGTGGTG GCAAACCCGGCAACAAGA-3'; *Ucp1 C253A* forward 5'-CCCAAGC GTACCAAGCGCTGCGATGTCCATGTAC-3', reverse 5'-GTACATGGAC ATCGCAGCGCTTGGTACGCTTGGG-3'; *Ucp1 C287A* forward 5'-GGAAC GTCATCATGTTTGTGGCCTTTGAACAGCTGAAAAAAG-3', reverse 5'-CTTTTTCAGCTGTTCAAAGGCCACAAACATGATGACGTTCC-3'; *Ucp1 C304A* forward 5'-CAGACAGACAGTGGATGCTACCACATAAGGATCC-3', reverse 5'-GGATCCTTATGTGGTAGCATCCACTGTCTGTCTG-3'.

Production of adenovirus and transduction of cysteine-null UCP1 mutants. pAd/CMV/V5-DEST/Ucp1 was linearized with PacI and transfected (3 μ g) into 293A cells with lipofectamine 2000 (Invitrogen). Crude adenovirus was generated according to the manufacturer's instructions (Invitrogen; V493-20). Crude adenovirus was amplified by infecting 293A cells, and purified using the Fast Trap Adenovirus Purification and Concentration Kit (EMD Millipore). Virus was quantified by examining viral DNA. Briefly, viral particles were treated with Proteinase K and DNA was isolated with phenol and chloroform/isoamylalcohol (24:1). Preliminary experiments with titrations of viral transductions in Ucp1^{-/-} adipocytes were used to determine the amount of virus yielding a Ucp1 messenger RNA (mRNA) and protein level similar to the level detected from Ucp1^{+/+} adipocytes. For subsequent experiments, primary brown adipocytes were transduced with purified adenovirus in the evening of day 3 after differentiation with medium replacement the following morning. Adipocytes were used for experiments on day 5 after differentiation.

Structural modelling of UCP1. A comparative model of UCP1 was built by using the structure of the bovine AAC¹⁹. This structure corresponds to the 'c-state' of the carrier—open to the mitochondrial inner membrane. The protein sequence of human UCP1 was taken from UniProt. To align the AAC and UCP1 sequences, MUSCLE⁴⁵ and manual editing in Jalview⁴⁶ were used. To improve the quality of the comparative models, the alignments were edited to remove the N- and C-terminal residues of the UCP1 sequences that did not align with resolved residues in the AAC structure, and to place gaps in the UCP1 sequences so as to minimize the distance between these residues in the initial target structure. Fifty comparative models of human UCP1 were built from the AAC structure and the sequence alignment by using MODELLER. The structure with the lowest

MODELLER energy score was taken as the best representative structure. The cardiolipin molecules of the AAC were added to the modelled UCP1 structure by aligning the two structures, and copying the lipid molecules^{21,22,47}. This structure was examined and figures produced by using the PyMOL molecular visualization system (PyMOL Molecular Graphics System, version 1.4.1, Schrödinger).

Assessment of superoxide levels in primary brown adipocytes. ROS production was estimated by oxidation of DHE and ratiometric assessment as described previously³³. Cells were plated and differentiated onto 96-well plates suitable for fluorescence analysis. Before imaging, cell media was removed and replaced with imaging buffer (156 mM NaCl, 1.25 mM KH₂PO₄, 3 mM KCl, 2 mM MgCl₂, 10 mM HEPES, pH 7.4) supplemented with 1 mM sodium pyruvate. Cells were loaded with 5 μ M DHE (Invitrogen), which remained present throughout the time course. DHE was excited at 355 nm and the emitted signal was acquired at 460 nm. Oxidized DHE was excited at 544 nm and emission was acquired at 590 nm.

Assessment of mitochondrial membrane potential in permeabilized primary brown adipocytes. Mitochondrial membrane potential was measured in permeabilized cells using TMRM (Life Technologies) in dequench mode. In this mode, mitochondrial depolarization causes redistribution of a high concentration of signal quenched TMRM from mitochondria to the cytosol, such that the lower concentration results in dequenching and an increase in fluorescence⁴⁸. Cells were pre-loaded at room temperature with imaging buffer containing 1 μ M TMRM. TMRM fluorescence was excited at 544 nm and emission was collected at 590 nm.

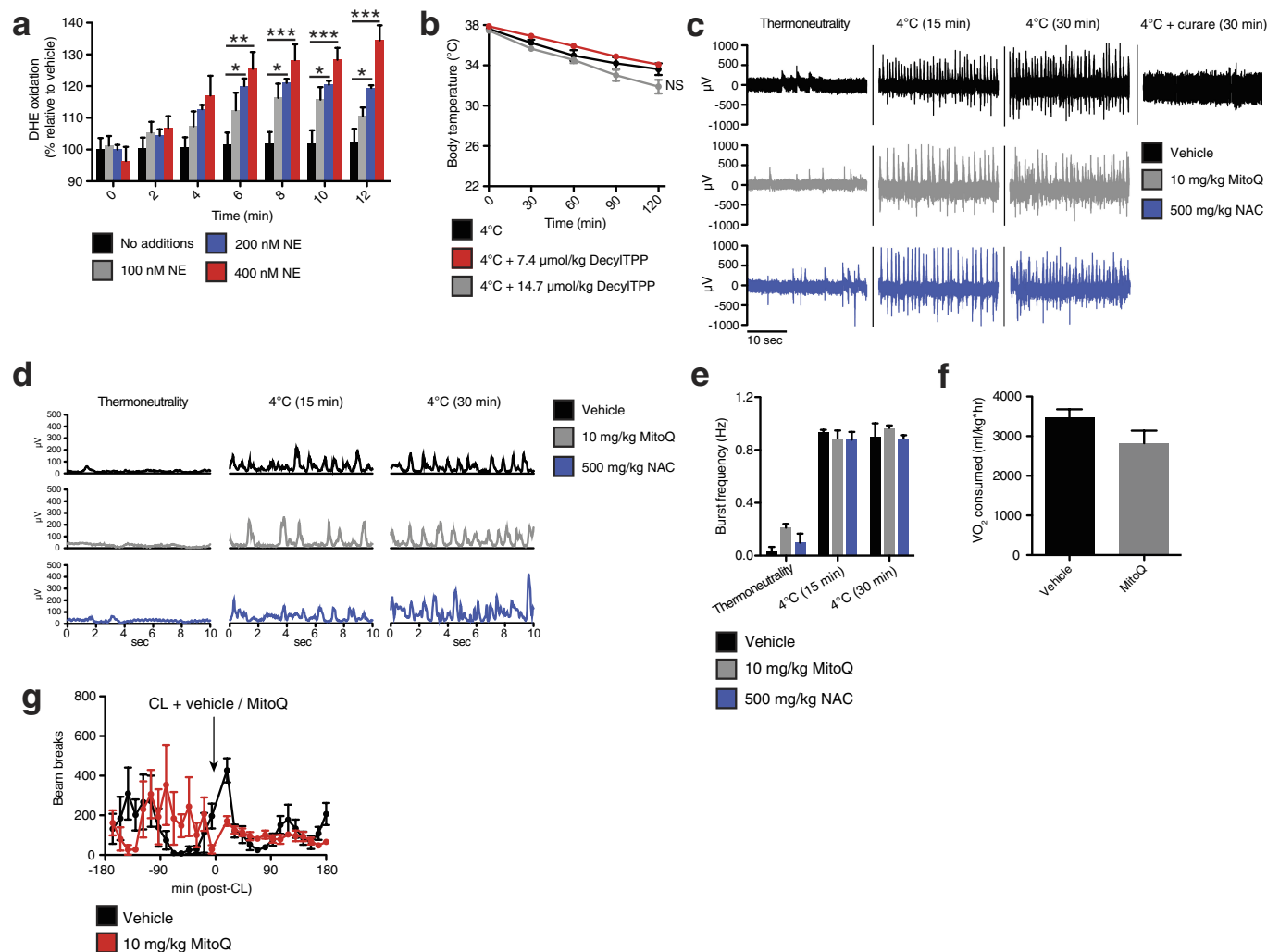
Quantitative RT-PCR. Total RNA was extracted from frozen tissue using TRIzol (Invitrogen), purified with RNeasy Mini spin columns (QIAGEN) and reverse transcribed using a High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). The resultant complementary DNA (cDNA) was analysed by quantitative PCR with reverse transcription (qRT-PCR). Briefly, 20 ng cDNA and 150 nmol of each primer were mixed with SYBR GreenER qPCR SuperMix (Applied Biosystems). Reactions were performed in a 384-well format using an ABI PRISM 7900HT real time PCR system (Applied Biosystems). Relative mRNA levels were calculated using the comparative CT method and normalized to cyclophilin mRNA. The following primers were used in these studies: *Cyclophilin* forward 5'-GGAGATGGCAGGAGGAA-3', reverse 5'-GCCCCGTAGTGCTTCAGCTT-3'; *Ucp1* forward 5'-ACTGCCACACCTCAGTCATT-3', reverse 5'-CTTTGCCTCACTCAGGATTGG-3'; *Dio2* forward 5'-CAGTGTGGTGCACGTCTCCAATC-3', reverse 5'-TGAACCAAAGTTGACCACCAG-3'; *Pgc1 α* forward 5'-CCCTGCCATTGTTAAGACC-3', reverse 5'-TGCTGCTGTCTCTGTTTC-3'; *PPAR- γ* forward 5'-TGAAAGAAGCGGTGAACCACTG-3', reverse 5'-TGGCATCTCTGTGTCAACCATG-3'; *Pgc1 β* forward 5'-CTGACGTGGACGAGCTTTCA-3', reverse 5'-CGTCCTTCAGAGCGTCAGAG-3'; *Nrf2* forward 5'-CCAGCTACTCCCAGGTTGCC-3', reverse 5'-GGGATATCCAGGGCAAGCGA-3'; *Ap2* 5'-AAGGTGAAGAGCATCATAACCCT-3', reverse 5'-TCACGCCTTTCATAACACATTCC-3'.

Glycerol release. Adipocytes were incubated in respiration medium absent BSA and treated with indicated concentrations of noradrenaline for 2 h before collection of medium and quantification of glycerol using free glycerol reagent (Sigma-Aldrich) relative to glycerol standard and normalized to protein content.

Immunoblot analysis. Immunodetection after SDS-PAGE used the following antibodies: UCP1 (Abcam ab10983), Prx3 (Abcam ab16751), Dimedone (Millipore 07-2139), Vinculin (Sigma V9264), ATP5A and NDUFB8 (Abcam ab110413), ATGL (CST 2138), ATGL pS406 (Abcam ab135093), HSL (CST 4107), HSL pS660 (CST 4126), pPKA substrate (CST 9624 s), PPAR- γ (CST 2435S).

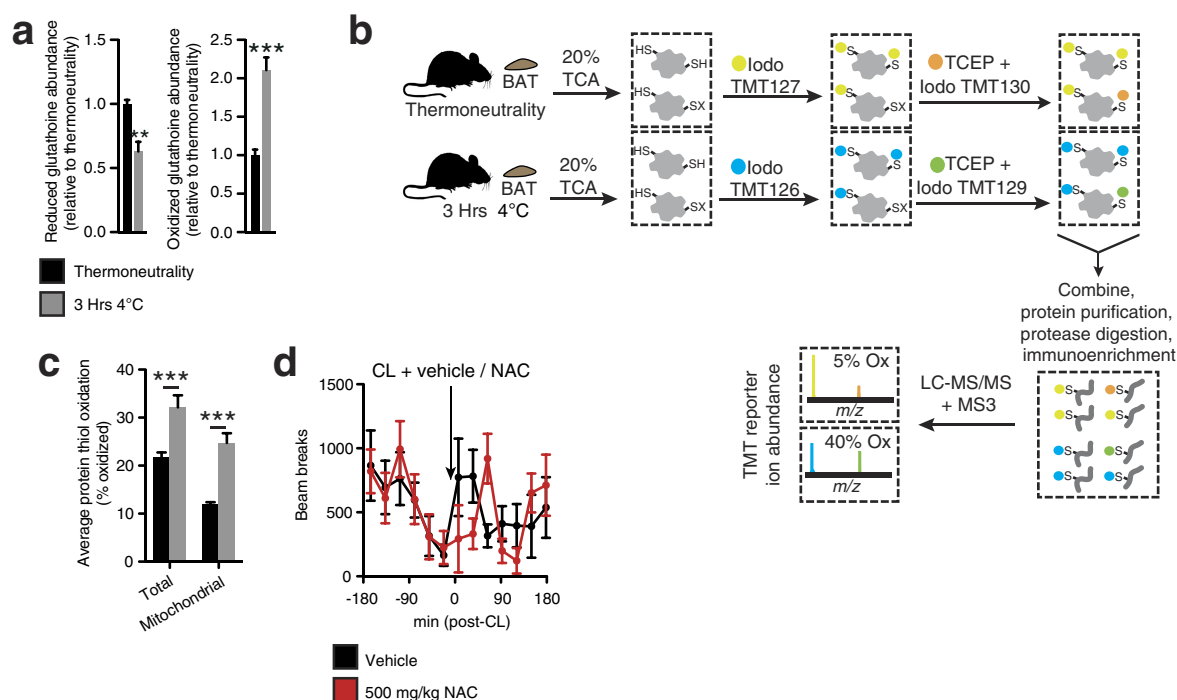
Statistical analyses. Data were expressed as mean \pm s.e.m. and *P* values were calculated using two-tailed Student's *t*-test for pairwise comparisons, one-way ANOVA for multiple comparisons, and two-way ANOVA for multiple comparisons involving two independent variables. ANOVA analyses were subjected to Bonferroni's post hoc test. Sample sizes were determined on the basis of previous experiments using similar methodologies. For *in vivo* studies, mice were randomly assigned to treatment groups. Mass spectrometric analyses were blinded to experimental conditions.

31. Gospodarska, E., Nowalis, P. & Kozak, L. P. Mitochondrial turnover: a phenotype distinguishing brown adipocytes from interscapular brown adipose tissue and white adipose tissue. *J. Biol. Chem.* **290**, 8243–8255 (2015).
32. Hodges, M. R. *et al.* Defects in breathing and thermoregulation in mice with near-complete absence of central serotonin neurons. *J. Neurosci.* **28**, 2495–2505 (2008).
33. Chouchani, E. T. *et al.* Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. *Nature* **515**, 431–435 (2014).
34. Srere, P. A. Controls of citrate synthase activity. *Life Sci.* **15**, 1695–1710 (1974).
35. Hurd, T. R. *et al.* Inactivation of pyruvate dehydrogenase kinase 2 by mitochondrial reactive oxygen species. *J. Biol. Chem.* **287**, 35153–35160 (2012).
36. Townsend, M. K. *et al.* Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin. Chem.* **59**, 1657–1667 (2013).
37. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2008).
38. Klomsiri, C. *et al.* Use of dimedone-based chemical probes for sulfenic acid detection evaluation of conditions affecting probe incorporation into redox-sensitive proteins. *Methods Enzymol.* **473**, 77–94 (2010).
39. Kazak, L. *et al.* A creatine-driven substrate cycle enhances energy expenditure and thermogenesis in beige fat. *Cell* **163**, 643–655 (2015).
40. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
41. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4**, 207–214 (2007).
42. Huttlir, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
43. McAlister, G. C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
44. Kir, S. *et al.* Tumour-derived PTH-related protein triggers adipose tissue browning and cancer cachexia. *Nature* **513**, 100–104 (2014).
45. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
46. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
47. Crichton, P. G. *et al.* Trends in thermostability provide information on the nature of substrate, inhibitor, and lipid interactions with mitochondrial carriers. *J. Biol. Chem.* **290**, 8206–8217 (2015).
48. Davidson, S. M., Yellon, D. & Duchen, M. R. Assessing mitochondrial potential, calcium, and redox state in isolated mammalian cells using confocal microscopy. *Methods Mol. Biol.* **372**, 421–430 (2007).
49. Held, J. M. *et al.* Targeted quantitation of site-specific cysteine oxidation in endogenous proteins using a differential alkylation and multiple reaction monitoring mass spectrometry approach. *Mol. Cell. Proteom.* **9**, 1400–1410 (2010).
50. Reisz, J. A., Bechtold, E., King, S. B., Poole, L. B. & Furdul, C. M. Thiol-blocking electrophiles interfere with labeling and detection of protein sulfenic acids. *FEBS J.* **280**, 6150–6161 (2013).



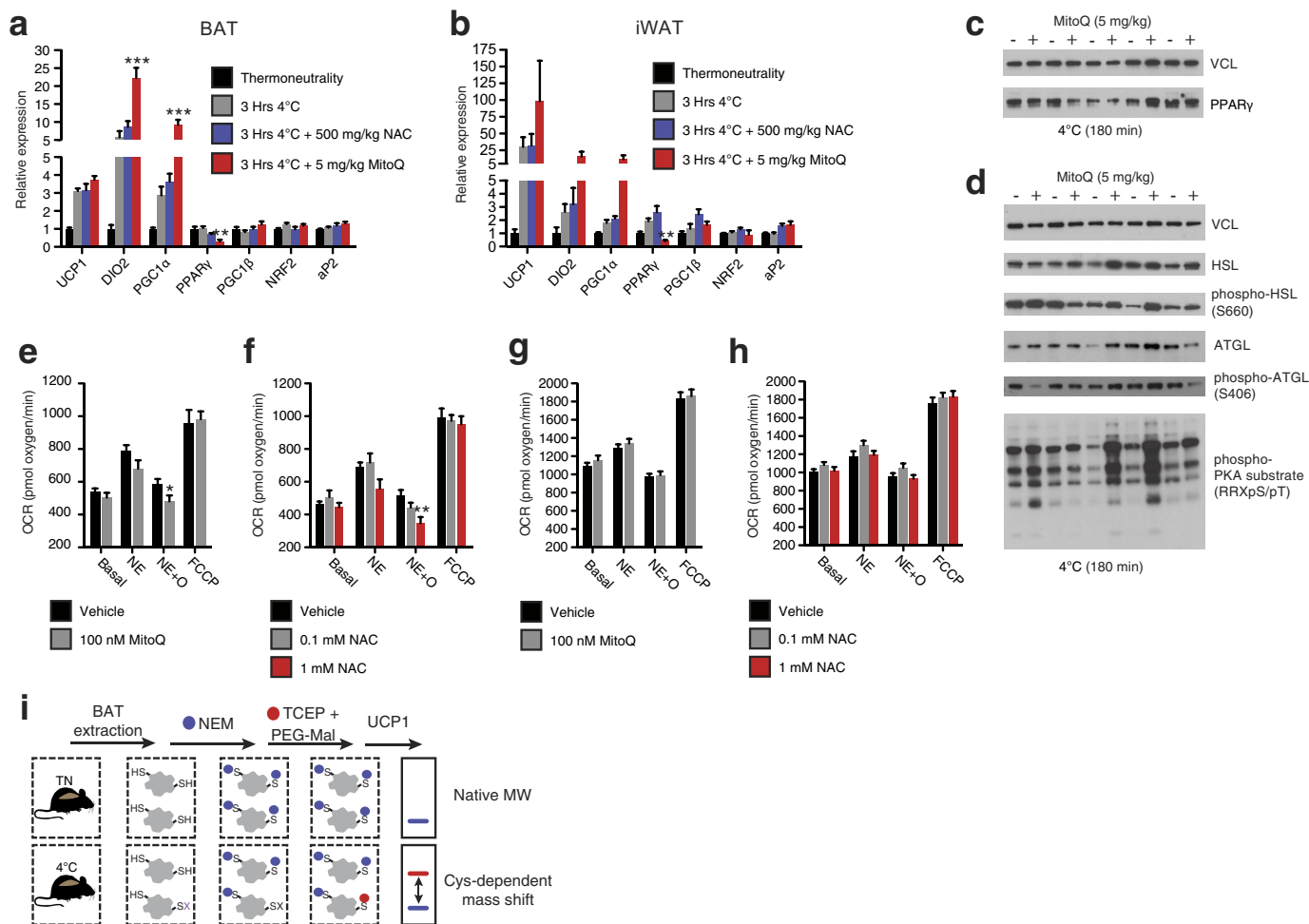
Extended Data Figure 1 | Assessing superoxide in brown adipocytes and effect of mitochondria-targeted compounds on shivering, body temperature, and movement. **a**, Noradrenaline stimulates superoxide-dependent oxidation of DHE in primary brown adipocytes ($n = 5$). **b**, Effect of i.p. decyl-TPP on core body temperature after acute cold exposure ($n = 10$). **c**, **d**, Representative (**c**) raw and (**d**) root mean square mouse EMG traces at thermoneutrality and after acute cold exposure \pm MitoQ, NAC, or curare (0.3 mg kg^{-1}). **e**, Quantification of muscle burst frequency as determined by EMG at thermoneutrality and

after acute cold exposure \pm MitoQ or NAC ($n = 3$). **f**, Absolute oxygen consumed immediately before acute CL treatment described in Fig. 1f ($n = 5$). **g**, Effect of i.p. CL \pm MitoQ on movement as assessed by number of beam breaks ($n = 8$). NS, not significant. Data are mean \pm s.e.m. of at least three replicates. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (two-tailed Student's t -test for pairwise comparisons, one-way ANOVA for multiple comparisons, two-way ANOVA for multiple comparisons involving two independent variables).



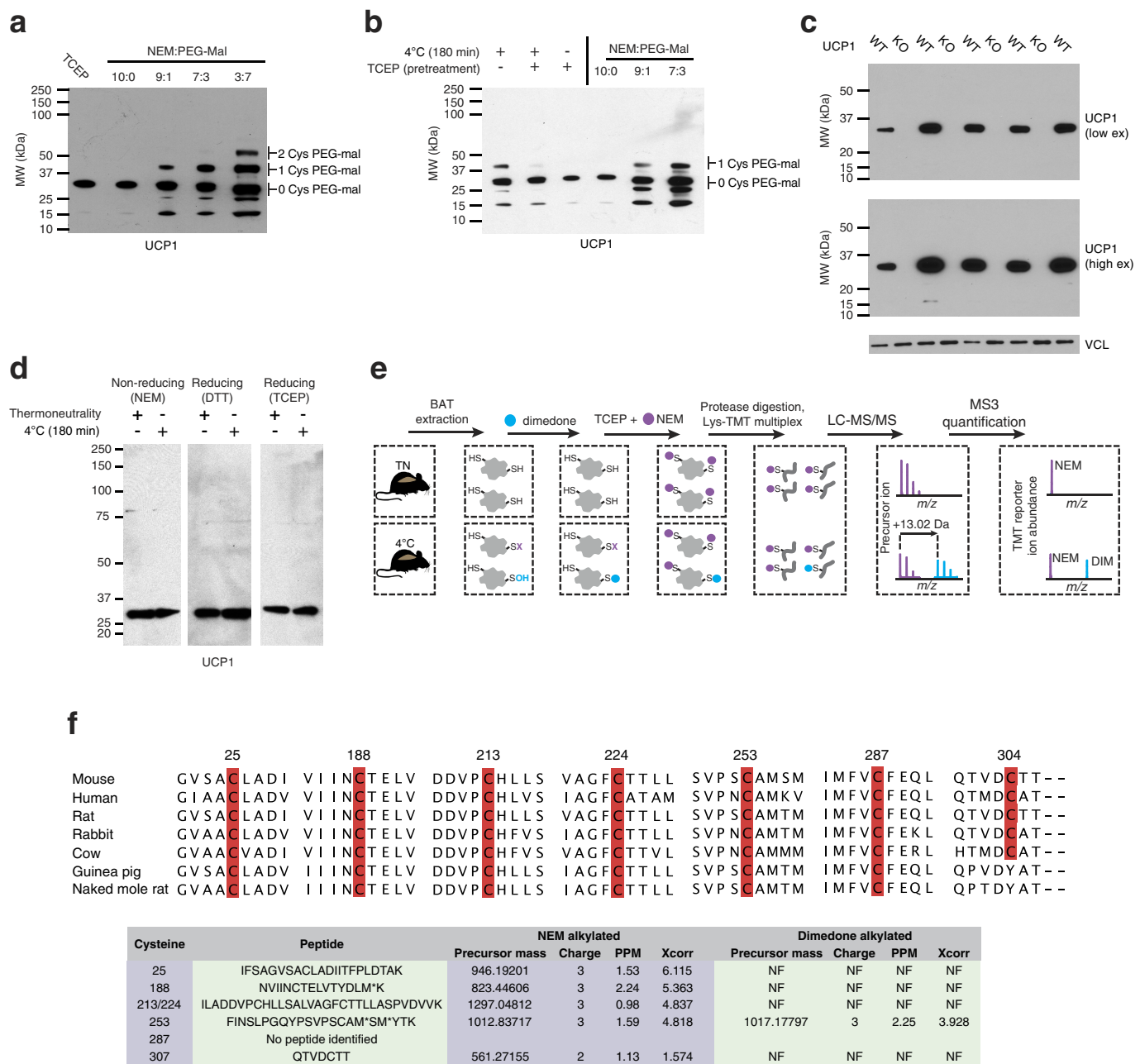
Extended Data Figure 2 | Assessing thiol redox status *in vivo* and the effect of NAC on movement. **a**, Mass spectrometric quantification of BAT-reduced and -oxidized glutathione at thermoneutrality and after acute cold exposure ($n = 5$). **b**, Scheme for quantitative assessment of protein thiol redox status by ratiometric labelling of BAT protein cysteine thiols. *In vivo* BAT thiol status is stabilized by protein precipitation in 20% TCA⁴⁹. Unmodified cysteine thiols are labelled with 'light' iodoTMT tags (126, 127, 128). After removal of light iodoTMT, reversibly modified protein thiols are reduced with TCEP in the presence of 'heavy' iodoTMT (129, 130, 131). Samples are combined and subjected to trypsin digestion.

Ratiometric assessment of iodoTMT labelled peptides provides a quantitative profile of overall protein cysteine thiol redox status. **c**, Average percentage oxidation status of total BAT and BAT mitochondrial protein thiols at thermoneutrality and after acute cold exposure ($n = 3$). **d**, Effect of i.p. NAC on movement as assessed by number of beam breaks (vehicle, $n = 11$; NAC, $n = 7$). Data are mean \pm s.e.m. of at least three replicates. $*P < 0.05$, $***P < 0.001$ (two-tailed Student's *t*-test for pairwise comparisons, one-way ANOVA for multiple comparisons, two-way ANOVA for multiple comparisons involving two independent variables).



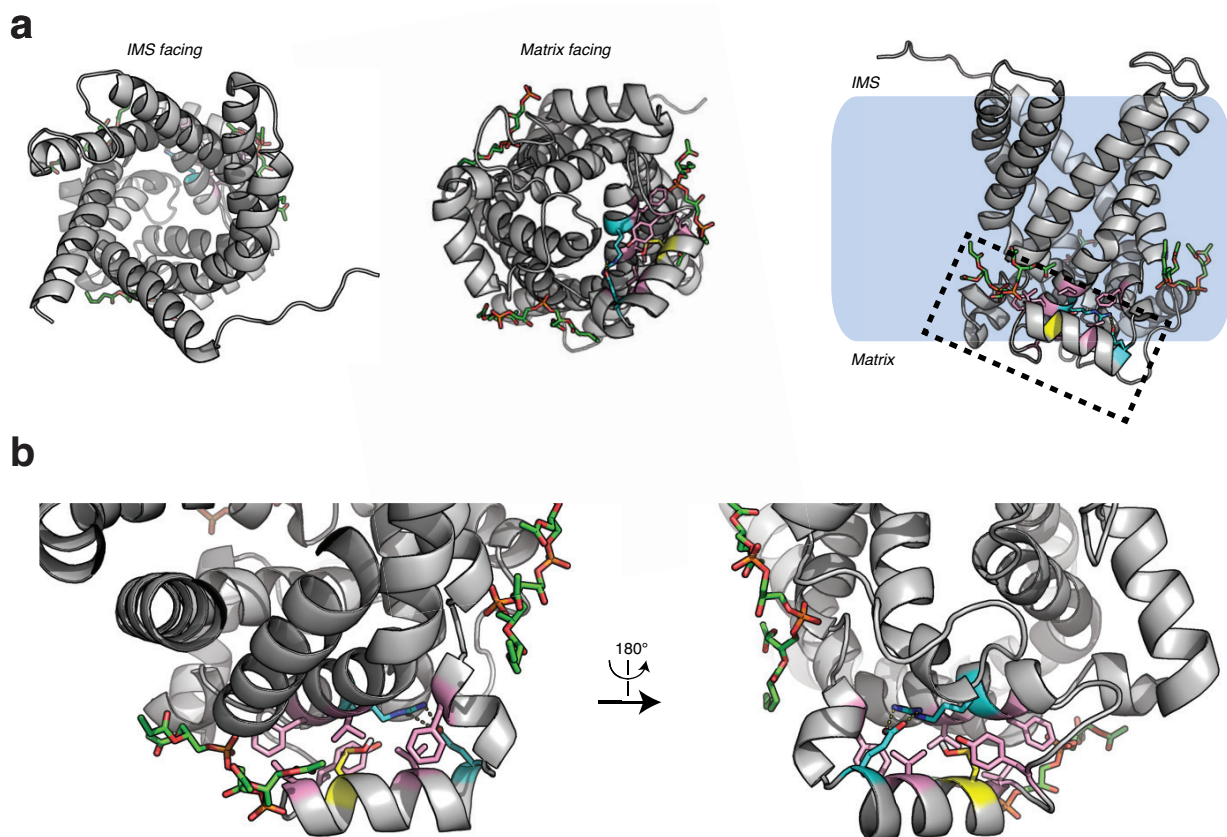
Extended Data Figure 3 | Assessing thermogenic gene expression, adrenergic response parameters, and strategy for determination of UCP1 cysteine thiol redox status. **a, b**, Quantitative polymerase chain reaction (qPCR) analysis of mRNA expression of selected **(a)** BAT and **(b)** inguinal white adipose tissue (iWAT) genes \pm cold exposure, \pm MitoQ or NAC ($n = 5$). **c, d**, Immunoblot analysis of **(c)** PPAR- γ protein expression levels and **(d)** lipolytic phosphorylation cascade activation in BAT after cold exposure \pm MitoQ. **e, f**, Raw OCR of primary brown adipocytes under basal conditions and after noradrenaline stimulation + oligomycin to determine leak respiration \pm **(e)** MitoQ ($n = 10$) or **(f)** NAC (vehicle and 1 mM NAC, $n = 8$; 0.1 mM NAC, $n = 7$). **g, h**, OCR of primary brown

adipocytes lacking UCP1 under basal conditions and after noradrenaline stimulation + oligomycin \pm **(g)** MitoQ ($n = 10$) and **(h)** NAC ($n = 10$). **i**, Cys-redox mass shift strategy. After *in vivo* interventions, mouse BAT is excised and unmodified protein thiols are labelled with NEM, after which reversibly oxidized thiols are chemically reduced and labelled with polyethyleneglycol maleimide (PEG-Mal) allowing for determination of cysteine oxidation status on UCP1. Data are mean \pm s.e.m. of at least five replicates. * $P < 0.05$, ** $P < 0.01$ (two-tailed Student's *t*-test for pairwise comparisons, one-way ANOVA for multiple comparisons, two-way ANOVA for multiple comparisons involving two independent variables).



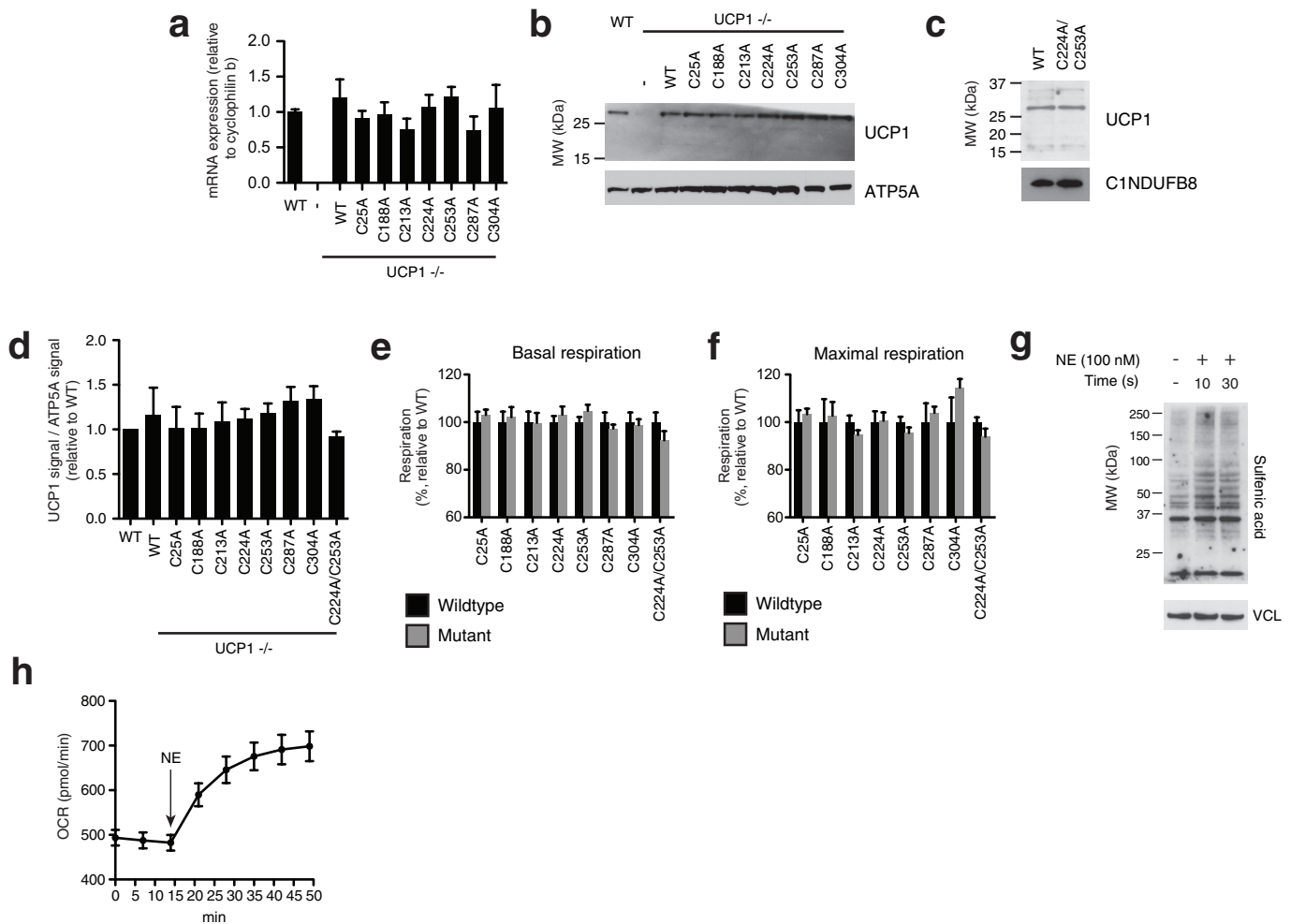
Extended Data Figure 4 | Assessing UCP1 reversible cysteine oxidation status *in vivo* by immunoblot and mass spectrometry. **a**, Calibration of UCP1 cysteine gel shift immunoblot. Calibration of cysteine-dependent shifts by incubation of BAT protein with TCEP and different ratios of NEM and PEG-Mal indicates that a single PEG-mal labelling event shifts UCP1 by ~10 kDa above the native molecular mass. **b**, Calibration of UCP1 cysteine oxidation status indicates that the gel shift observed upon cold exposure (lane 1) is cysteine dependent, as TCEP pretreatment results in a loss of the shift (lane 2). In addition, the cysteine-dependent mass shift is due to a single oxidation event as determined by including the calibrating markers (lanes 4–6). **c**, Calibration of specificity of UCP1 antibody in BAT. **d**, Reducing and non-reducing SDS-PAGE analysis of UCP1 to monitor cysteine-dependent and -independent inter-protein interactions during acute cold exposure. **e**, Scheme for identification of sulfenylated cysteines on UCP1 by dimedone labelling and mass spectrometry. After acute cold exposure, BAT protein thiols are differentially alkylated with dimedone to selectively label sulfenylated

thiols and NEM to label non-sulfenylated thiols. Samples are subjected to trypsin digestion, followed by Lys-TMT labelling, and MS quantification of UCP1 cysteine containing peptides in their dimedone and NEM alkylated forms. Two technical points should be noted in this strategy when interpreting relative quantitation of NEM and dimedone-alkylated peptides. First, these differently alkylated peptides may not necessarily ionize with the same efficiency. Second, NEM is reported to react with sulfenic acids albeit less efficiently than with free thiols⁵⁰, which should be factored in when considering the order of addition of dimedone/ NEM and potential underestimation of sulfenylation status. **f**, Top: amino-acid sequence alignment of UCP1 proteins highlighting the candidate cysteine residues contained within the mouse protein and their level of conservation across various species. Bottom: summary of MS determination of UCP1 cysteine sulfenylation status. Six of seven UCP1 cysteines were identified, with all but one being identified exclusively in the unmodified (NEM-alkylated state). Cys253 is identified as dimedone labelled, indicating that it is a site for sulfenylation.



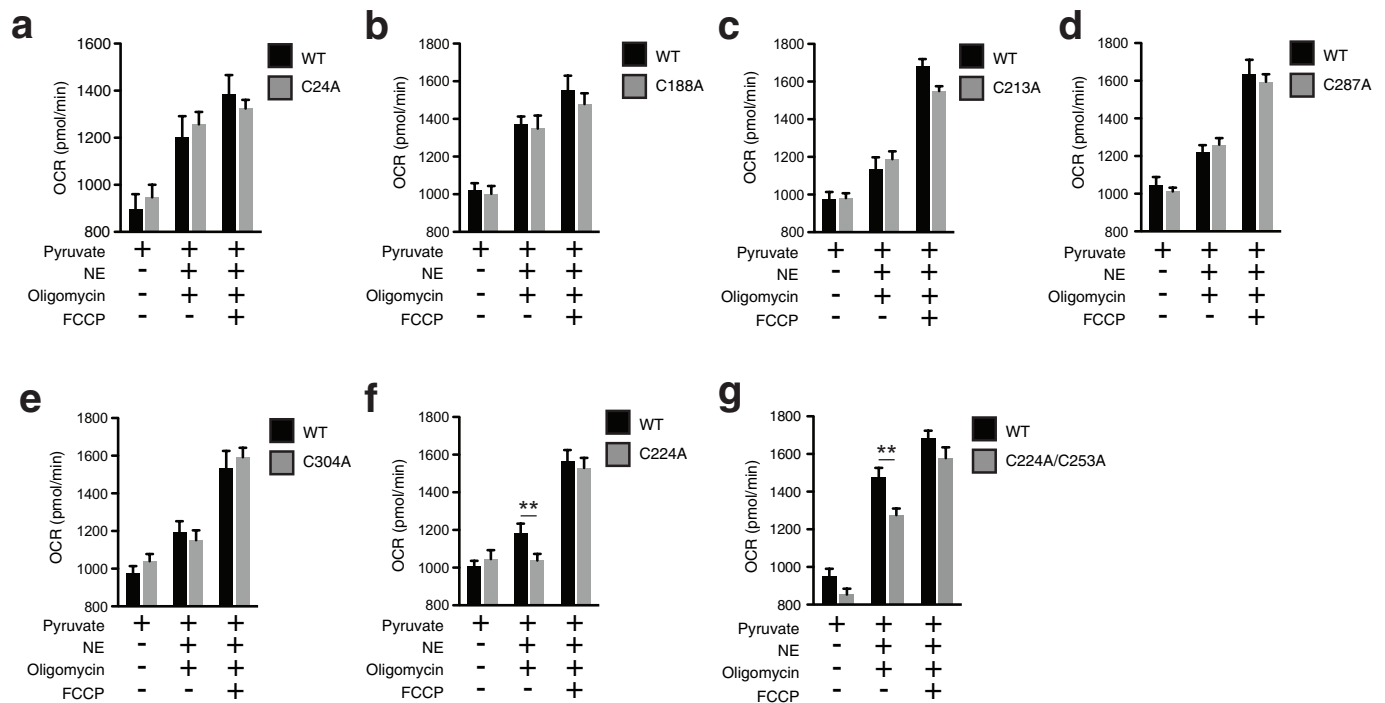
Extended Data Figure 5 | Structure of human UCP1 modelled on the AAC crystal structure including bound cardiolipin and sulfenylation of Cys253. a, Entire UCP1 modelled structure including bound cardiolipin (green), and Cys253 in the oxidized sulfenic acid form. **b,** UCP1 region containing Cys253 in the oxidized sulfenic acid form. Cys253 localizes to a hydrophobic pocket between two matrix facing helices. Hydrophobic

residues (pink) surround the Cys253 thiol, and a hydrogen bond between Arg238 and Glu261 (aqua) is proximal. These residues that stabilize interaction between the matrix facing helices are probably important for stabilization of the purine bound 'c-state' of the carrier. Two separate cardiolipin (green) binding domains are localized proximal to Cys253 within the UCP1 modelled structure.



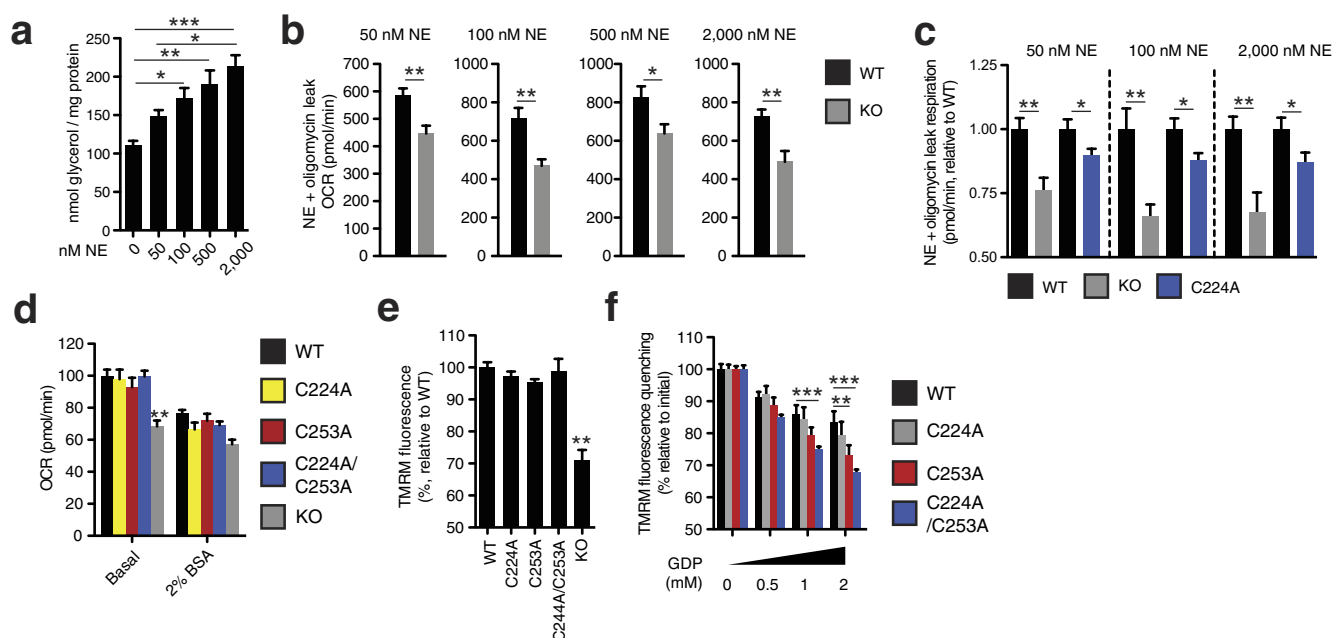
Extended Data Figure 6 | Assessing transduced UCP1 constructs, OCR, and sulfenylation in brown adipocytes. **a**, Quantitative PCR analysis of UCP1 mRNA in WT, and UCP1 $^{-/-}$ brown adipocytes transduced with WT and cysteine null UCP1 mutants ($n = 4$). **b**, Immunoblot of UCP1 protein in WT and UCP1 $^{-/-}$ brown adipocytes transduced with WT and cysteine null UCP1 mutants. **c**, Immunoblot analysis of UCP1 protein in UCP1 $^{-/-}$ brown adipocytes transduced with C224A/C253A double mutant compared with WT. **d**, Densitometry analysis of transduced UCP1

forms relative to WT across separate transduction experiments ($n = 4$; C224A/C253A $n = 3$). **e**, **f**, Summary of basal (**e**) and maximal (**f**) OCR of primary brown adipocytes containing cysteine-null UCP1 mutants. Raw OCR values provided in Extended Data Fig. 7. **g**, Immunodetection of protein sulfenic acid levels in primary brown adipocytes in the seconds after treatment with 100 nM noradrenaline. **h**, Time course of brown adipocyte OCR after stimulation with 100 nM noradrenaline ($n = 12$). Data are mean \pm s.e.m. of at least three replicates.



Extended Data Figure 7 | Assessing OCR under basal and FCCP-stimulated maximal respiration, and after noradrenaline stimulation + oligomycin in UCP1^{-/-} primary brown adipocytes transduced with UCP1 cysteine null mutants. a–g, Raw basal, maximal, and UCP1-dependent OCR from representative experiments of WT and UCP1^{-/-} brown adipocytes transduced with (a) UCP1 C24A

(WT $n = 11$; C24A $n = 12$), (b) C188A ($n = 7$), (c) UCP1 C213A ($n = 19$), (d) UCP1 C287A (WT $n = 9$; C287A $n = 10$), (e) UCP1 C304A (WT $n = 7$; C304A $n = 10$), (f) UCP1 C224A (WT $n = 9$; C224A $n = 10$), (g) UCP1 C224A/C253A ($n = 8$). Data are mean \pm s.e.m. of at least seven replicates. ** $P < 0.01$ (two-tailed Student's t -test for pairwise comparisons).



Extended Data Figure 8 | Assessing UCP1-dependent respiration and uncoupling following increasing degrees of adrenergic stimulus.

a, Glycerol release from brown adipocytes as an index of adrenergic stimulus and lipolysis in response to increasing concentrations of noradrenaline ($n = 4$; 0 and 2,000 nM noradrenaline $n = 6$). **b**, Representative raw noradrenaline + oligomycin leak OCR values in WT and UCP1^{-/-} brown adipocytes after stimulation with a range of noradrenaline concentrations indicates that UCP1-dependent leak respiration is consistently ~25–35% of total leak OCR (50 nM $n = 9$; 100 nM $n = 7$; 500 nM $n = 8$; 2,000 nM $n = 6$). **c**, Assessment of UCP1-dependent leak respiration after stimulation by various concentrations of noradrenaline + oligomycin. Comparison of WT and UCP1^{-/-} OCR (replotted data from Fig. 4 for comparison) indicates that UCP1-dependent respiration is consistently ~25–35% of leak respiration.

Comparison of UCP1 WT and C224A indicates that the degree of UCP1 inhibition by C224A is relatively stable across various noradrenaline concentrations ($n = 8$; C224A 100 nM noradrenaline $n = 9$, 2,000 nM noradrenaline $n = 14$). **d**, Plasma-membrane-permeabilized OCR of brown adipocytes ± endogenous fatty-acid depletion with BSA (WT $n = 30$; KO $n = 20$; C224A $n = 9$; C253A $n = 8$; C224A/C253A $n = 10$). **e**, Comparison of UCP1-dependent uncoupling absent purine nucleotide inhibition in plasma-membrane-permeabilized brown adipocytes ($n = 6$; C224A/C253A, KO $n = 4$). **f**, Comparison of UCP1-dependent uncoupling after titration of GDP in permeabilized adipocytes containing WT UCP1, UCP1 C224A, C253A, and C224A/C253A ($n = 6$; C224A/C253A $n = 12$). Data are mean ± s.e.m. of at least four replicates. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (two-tailed Student's t -test for pairwise comparisons, one-way ANOVA for multiple comparisons).

Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia

Ryu Okumura^{1,2}, Takashi Kurakawa¹, Takashi Nakano³, Hisako Kayama^{1,2}, Makoto Kinoshita^{1,2}, Daisuke Motooka⁴, Kazuyoshi Gotoh^{4,5}, Taishi Kimura¹, Naganori Kamiyama¹, Takashi Kusu¹, Yoshiyasu Ueda¹, Hong Wu³, Hideki Iijima⁶, Soumik Barman^{1,2}, Hideki Osawa⁷, Hiroshi Matsuno⁷, Junichi Nishimura⁷, Yusuke Ohba⁸, Shota Nakamura⁴, Tetsuya Iida^{4,9}, Masahiro Yamamoto¹⁰, Eiji Umemoto^{1,2}, Koichi Sano³ & Kiyoshi Takeda^{1,2}

Colonic epithelial cells are covered by thick inner and outer mucus layers^{1,2}. The inner mucus layer is free of commensal microbiota, which contributes to the maintenance of gut homeostasis^{3–6}. In the small intestine, molecules critical for prevention of bacterial invasion into epithelia such as Paneth-cell-derived anti-microbial peptides and regenerating islet-derived 3 (RegIII) family proteins have been identified^{7–11}. Although there are mucus layers providing physical barriers against the large number of microbiota present in the large intestine, the mechanisms that separate bacteria and colonic epithelia are not fully elucidated. Here we show that Ly6/PLAUR domain containing 8 (Lypd8) protein prevents flagellated microbiota invading the colonic epithelia in mice. Lypd8, selectively expressed in epithelial cells at the uppermost layer of the large intestinal gland, was secreted into the lumen and bound flagellated bacteria including *Proteus mirabilis*. In the absence of Lypd8, bacteria were present in the inner mucus layer and many flagellated bacteria invaded epithelia. *Lypd8*^{−/−} mice were highly sensitive to intestinal inflammation induced by dextran sulfate sodium (DSS). Antibiotic elimination of Gram-negative flagellated bacteria restored the bacterial-free state of the inner mucus layer and ameliorated DSS-induced intestinal inflammation in *Lypd8*^{−/−} mice. Lypd8 bound to flagella and suppressed motility of flagellated bacteria. Thus, Lypd8 mediates segregation of intestinal bacteria and epithelial cells in the colon to preserve intestinal homeostasis. The intestinal mucosa is protected from commensal microbes and pathogenic microorganisms by several types of barriers^{5,6}, including mucus layers produced by intestinal secretory cells¹. Compared with those in the small intestine, epithelial cells in the large intestine are covered by thick mucus composed of two layers: the inner firm mucus layer and outer loose mucus layer. Abundant numbers of commensal microbiota are present in the outer mucus layer, whereas none is present in the inner mucus layer^{3,4}. The firm inner mucus layer acts as a physical barrier, but it is possible that other mechanisms for maintaining the segregation of bacteria and colonic epithelia also exist. To address this issue, we searched for a gene that is exclusively expressed in the large intestine in a public database (<http://biogps.org/#goto=welcome>), and found that *Lypd8* was selectively expressed in epithelial cells of the mouse gastrointestinal tract, especially the caecum and large intestine (Fig. 1a, b). *Lypd8* was expressed in germ-free mice at a similar level to that in specific-pathogen-free (SPF) mice (Extended Data Fig. 1a). *Lypd8* was expressed at normal levels in *Myd88*^{−/−} and *Asc*^{−/−} (also known as *Pycard*) mice (Extended Data Fig. 1b), indicating

that *Lypd8* expression is not induced by microbiota-dependent inflammatory signals. We observed high *Lypd8* mRNA expression *in situ* in the uppermost layer of the colonic gland exposed to the lumen of the large intestine (Fig. 1c), and when introduced into a polarized human colorectal carcinoma Caco-2 cell line, Lypd8 was expressed at the apical surface (Extended Data Fig. 1c). Thus, Lypd8 is expressed on the apical surface of epithelial cells located at the uppermost layer of the colonic gland.

On the basis of amino acid length, the expected molecular weight of Lypd8 protein is about 25 kilodaltons (kDa). However, western blot analysis showed a band for recombinant Lypd8 protein that was greater than 100 kDa. Lypd8 carries 13 asparagine (Asn) residues that are potential N-glycosylation sites. We therefore treated recombinant Lypd8 protein with peptide-N-glycosidase F (PNGase F), an amidase that cleaves between the innermost N-acetylglucosamine and Asn residues of the N-linked glycoproteins. Treatment with PNGase F substantially reduced the molecular mass of Lypd8 proteins (Fig. 1d). Treatment with sialidase A gave a more diffuse band, as sialic acids were removed, and the addition of O-glycanase did not further alter the mobility, showing that Lypd8 did not contain any Galβ(1-3)GalNAc-O-glycans. Replacing all of the Asn residues with Asp resulted in the appearance of a 25 kDa Lypd8 protein (Extended Data Fig. 1d), indicating that Lypd8 is a highly N-glycosylated protein.

Lypd8 is a member of the Ly6/PLAUR family of proteins, many of which are glycosylphosphatidylinositol (GPI)-anchored^{12,13}. We treated a mouse colon cancer CMT93 cell line stably expressing Flag-tagged Lypd8 with phosphatidylinositol-specific phospholipase C (PI-PLC), which cleaves and releases GPI-anchored proteins from the cell membrane, and found that surface Lypd8 expression was severely reduced after PI-PLC treatment (Extended Data Fig. 1e). We then examined whether Lypd8 is secreted into the intestinal lumen using immunohistochemistry with a newly established anti-Lypd8 monoclonal antibody (mAb) and Carnoy's fixative, which preserves the mucus layer³ (Fig. 1e). Positive staining for Lypd8 was observed in epithelial cells at the uppermost layer of the colonic gland, and a positive signal was also observed in the intestinal lumen. In particular, an intense Lypd8 signal was observed at the interface of inner and outer mucus layers that are exposed to commensal bacteria. In addition, analyses using *Lypd8*^{venus} mice, in which modified yellow fluorescent protein is expressed under the control of the *Lypd8* promoter, showed that Lypd8-expressing cells also expressed *Villin*, but not *Muc2* or *Lgr5* (Extended Data Fig. 2). Thus, Lypd8 is a GPI-anchored highly glycosylated protein expressed in

¹Laboratory of Immune Regulation, Department of Microbiology and Immunology, Graduate School of Medicine, WPI Immunology Frontier Research Center, Osaka University, Suita, Osaka 565-0871, Japan. ²Core Research for Evolutional Science and Technology, Japan Agency for Medical Research and Development, Tokyo 100-0004, Japan. ³Department of Microbiology and Infection Control, Osaka Medical College, Takatsuki, Osaka 569-8686, Japan. ⁴Department of Infection Metagenomics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan. ⁵Department of Bacteriology, Okayama University Graduate School of Medicine, Okayama 700-8558, Japan. ⁶Department of Gastroenterology and Hepatology, Graduate School of Medicine, Osaka University, Osaka 565-0871, Japan. ⁷Department of Gastroenterological Surgery, Graduate School of Medicine, Osaka University, Osaka 565-0871, Japan. ⁸Department of Cell Physiology, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan. ⁹Department of Bacterial Infections, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan. ¹⁰Laboratory of Immunoparasitology, Research Institute for Microbial Diseases, WPI Immunology Frontier Research Center, Osaka University, Suita, Osaka 565-0871, Japan.

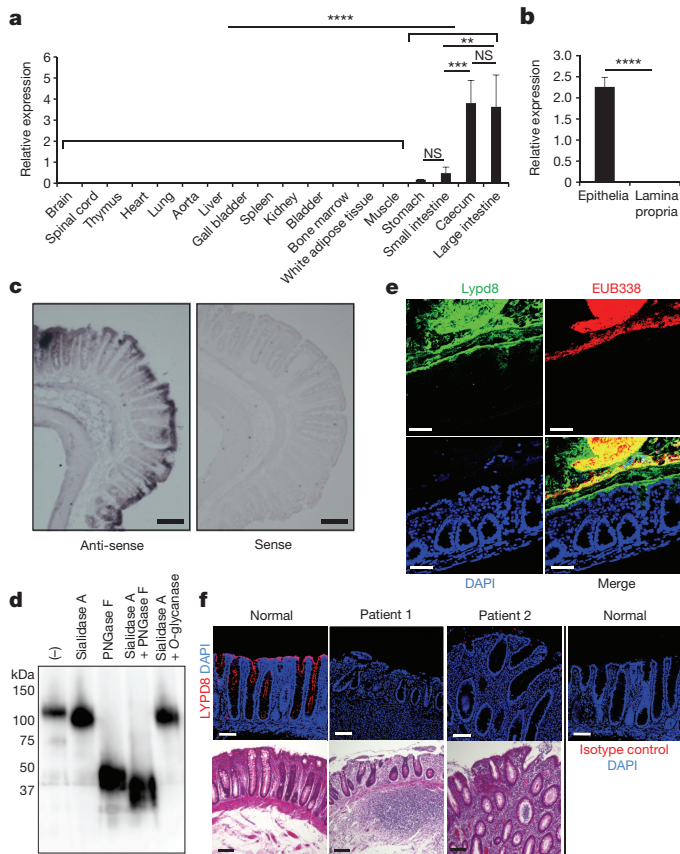


Figure 1 | Molecular characteristics of *Lypd8*. **a, b**, Quantitative reverse transcription (RT)–PCR of *Lypd8* expression in various organs (**a**), and in the epithelia and lamina propria of the colon (**b**). Values were normalized to *Gapdh* expression. Data represent the mean \pm s.d. ($n = 4$ per group). ** $P < 0.01$, *** $P < 0.005$, **** $P < 0.001$, NS, not significant. **c**, *In situ* hybridization of colon sections with anti-sense and sense *Lypd8* RNA probes. Scale bars, 50 μ m. **d**, Western blot of recombinant Flag-tagged *Lypd8* proteins untreated or treated with sialidase A, PNGase F or O-glycanase. **e**, Immunostaining with anti-*Lypd8* antibody (green), FISH using an EUB338 probe (red), and 4',6-diamidino-2-phenylindole (DAPI; blue) of Carnoy's fixed colon sections. Scale bar, 50 μ m. **f**, Immunostaining (upper panel) with anti-*Lypd8* antibody (red) and DAPI (blue) and H&E staining (lower panel) of human colon sections from the normal mucosa of a colon cancer patient and the inflamed mucosa of two ulcerative colitis patients. Scale bar, 100 μ m.

enterocytes located at the uppermost epithelial layer of the colon, which is cleaved and secreted into the colonic lumen. In healthy human intestinal sections, LYPD8 was preferentially and highly expressed on the epithelial layer exposed to the lumen (Fig. 1f); however, LYPD8 staining was severely decreased in patients with ulcerative colitis (Fig. 1f), indicating the possible involvement of this protein in its pathogenesis.

We next analysed the effect of the *Lypd8* deficiency (Extended Data Fig. 3). *Lypd8*^{−/−} mice were healthy when reared in an SPF facility in which bacterial contamination was strictly controlled. Histological analysis of haematoxylin and eosin (H&E)-stained colonic tissue showed a region that was free of luminal contents above the epithelial cell layer in wild-type mice, but was colonized by bacteria in *Lypd8*^{−/−} mice (Fig. 2a). We examined the presence of intestinal bacteria adjacent to the epithelial layer in the large intestine by fluorescence *in situ* hybridization (FISH) using a universal bacterial 16S ribosomal RNA (rRNA) probe (Fig. 2b, c). This area was free from bacteria in wild-type mice, but bacteria were detected in close proximity to the epithelial layers in *Lypd8*^{−/−} mice despite normal mucin production, as confirmed by anti-Muc2 antibody staining (Fig. 2d). Furthermore, there was no positive signal in the crypts of wild-type mice, but bacteria invaded the

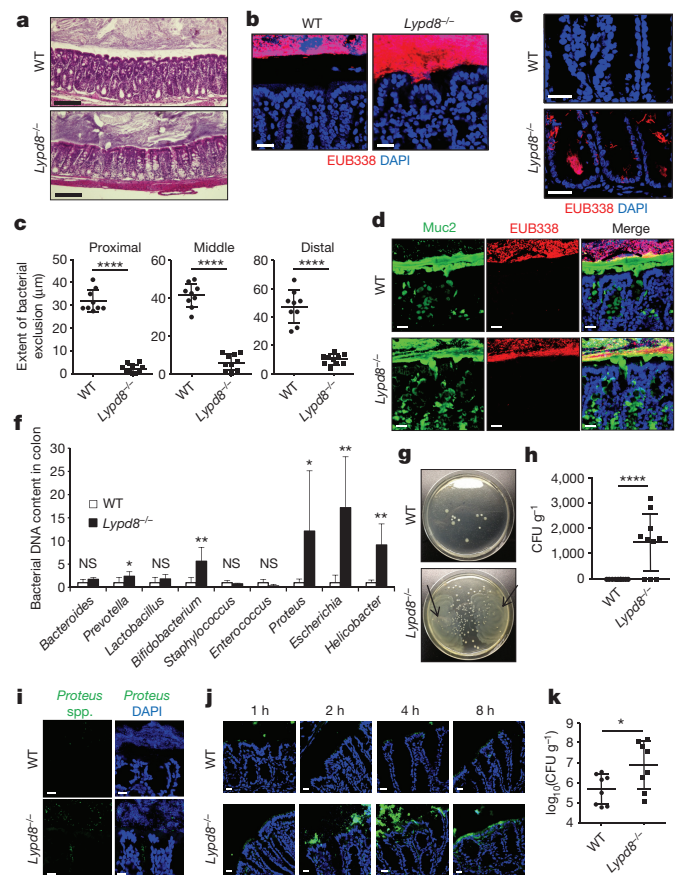


Figure 2 | Penetration of the colonic epithelia by flagellated bacteria in *Lypd8*^{−/−} mice. **a**, H&E staining of Carnoy's fixed colon sections. Scale bars, 100 μ m. **b**, FISH using EUB338 probe (red) and DAPI (blue) of the Carnoy's fixed colon. Scale bars, 20 μ m. **c**, Extent of bacterial exclusion from the tip of epithelia to luminal bacteria in proximal, middle, and distal colon. ($n = 9$ wild-type (WT) mice; $n = 10$ *Lypd8*^{−/−} mice). **d**, Immunostaining with anti-Muc2 antibody (green), EUB338 (red) and DAPI (blue) of Carnoy's fixed colon sections. Scale bars, 20 μ m. **e**, EUB338 (red) and DAPI (blue) in the crypt region of the colon. Scale bars, 20 μ m. **f**, Quantitative PCR of bacterial DNA in 1 mg of colon tissue, using primers specific for each bacterial genus. Data show each bacterial DNA amount compared to wild-type mice group ($n = 6$ per group). **g**, Bacterium forming the swarming colony (arrows). **h**, Colony-forming units (CFU) of *P. mirabilis* in colon tissues. **i**, FISH using *Proteus*-specific probe (green) and DAPI (blue) in the colon. **j**, Immunohistochemistry analysis of colon sections at the indicated time points after inoculation of CFSE-labelled *P. mirabilis*. *P. mirabilis* (green) and DAPI (blue). The representative of three independent experiments is shown. Scale bars, 20 μ m. **k**, CFU of *P. mirabilis* in colon tissues at 8 h after the inoculation. Data are mean \pm s.d. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.001$.

crypts of *Lypd8*^{−/−} mice (Fig. 2e). Thus, intestinal bacteria penetrated the inner mucus layer and further into the crypts in the large intestine of *Lypd8*^{−/−} mice.

The bacterial composition of colonic luminal contents analysed by 16S rRNA gene sequencing was not altered in *Lypd8*^{−/−} mice (Extended Data Fig. 4a). Therefore, we isolated DNA from the luminal content and colonic tissues and analysed for the presence of bacteria by PCR using universal primers for bacterial 16S rRNA (Extended Data Fig. 4b). Bacterial DNA was similarly amplified in luminal contents from wild-type and *Lypd8*^{−/−} mice. In wild-type colonic tissues, minimal DNA was detected. In contrast, bacterial DNA was substantially amplified in colonic tissues from *Lypd8*^{−/−} mice. We then identified the bacterial species invading *Lypd8*^{−/−} epithelial cell layers using bacterial genus-specific primers (Fig. 2f). The numbers of several bacterial genera were increased in *Lypd8*^{−/−} colonic tissues, particularly *Proteus*,

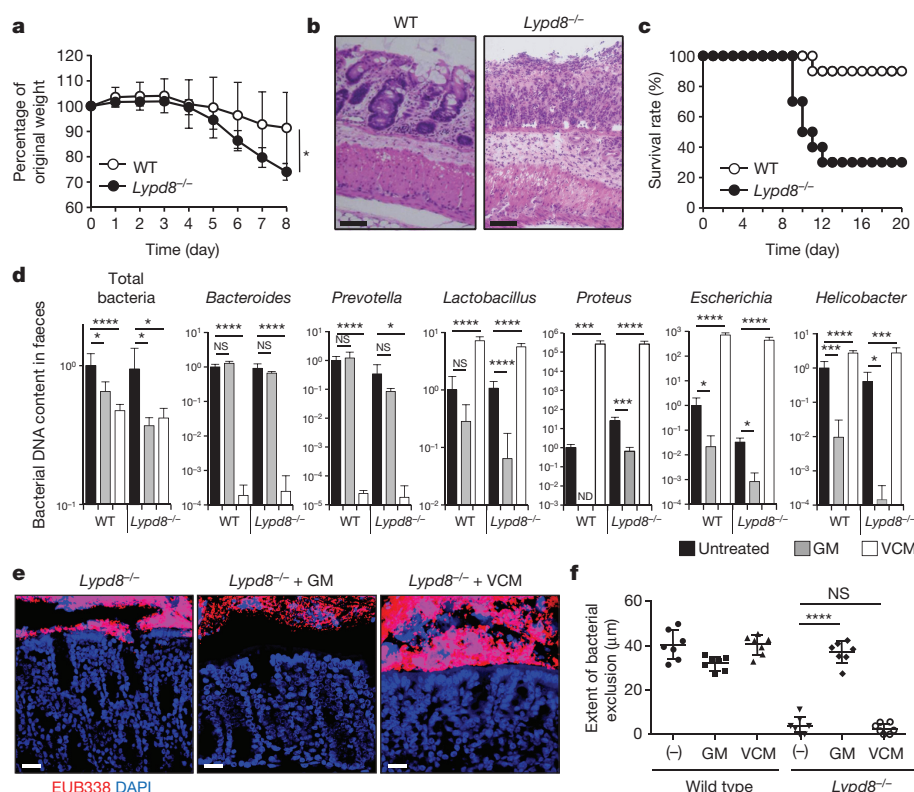


Figure 3 | High sensitivity to intestinal inflammation in *Lypd8*^{-/-} mice.

a–c, Percentage of weight change after 2% DSS administration of wild-type ($n = 10$) and *Lypd8*^{-/-} mice ($n = 10$) (**a**). H&E staining of colon tissues at day 8 of DSS administration (**b**). Scale bars, 50 μ m. Survival rate (**c**). **d**, Quantitative PCR of bacterial DNA isolated from 1 mg of faeces from untreated, gentamicin (GM)-treated and vancomycin (VCM)-treated

mice. Data show the bacterial DNA amounts compared to untreated wild-type mice group ($n = 5$ per group). ND, not detected. **e**, FISH using EUB338 probe (red) and DAPI (blue) in the colon of *Lypd8*^{-/-} mice untreated or orally treated with GM or VCM. Scale bars, 20 μ m. **f**, Extent of bacterial exclusion from the tip of epithelia to luminal bacteria ($n = 7$ per group). Data are mean \pm s.d. * $P < 0.05$, *** $P < 0.005$, **** $P < 0.001$.

Helicobacter and *Escherichia*. Next, homogenized colonic tissues from wild-type and *Lypd8*^{-/-} mice were cultured on agar plates (Fig. 2g). Several bacterial colonies grew from wild-type and *Lypd8*^{-/-} colonic tissues. Among them, unique colonies with a swarming migration pattern on agar were present in the *Lypd8*^{-/-} colonic tissue homogenates. DNA sequence analysis revealed that the bacterium forming the unusual swarming colony derived from *Lypd8*^{-/-} tissues was *Proteus mirabilis* (Fig. 2h). We analysed the presence of *Proteus* spp. in colonic tissues by FISH using a *Proteus*-specific 16S rRNA probe (Fig. 2i). *Proteus* spp. were detected in the luminal regions of wild-type mice, but slightly increased numbers of *Proteus* spp. were observed in the luminal regions of *Lypd8*^{-/-} mice. Furthermore, *Proteus* spp. were present just above the epithelial cell layer and in the crypts of *Lypd8*^{-/-} mice. We administered a transanal injection of carboxyfluorescein succinimidyl ester (CFSE)-labelled *P. mirabilis* to wild-type and *Lypd8*^{-/-} mice, and analysed the colonic tissues by immunohistochemistry at various time points (Fig. 2j). In wild-type mice, *P. mirabilis* was not observed adjacent to epithelial cell layers, but was in close proximity to the epithelial cell layers of *Lypd8*^{-/-} mice 2 h after administration. Furthermore, *P. mirabilis* was present in closer proximity to the epithelial layers of *Lypd8*^{-/-} mice at 8 h (Fig. 2k). Despite the penetration of the epithelial surface by commensal bacteria, an inflammatory response (as determined by inflammatory cytokine expression) was not evident in the colon of *Lypd8*^{-/-} mice reared in an SPF facility (Extended Data Fig. 5). Thus, intestinal bacteria such as *Proteus*, *Helicobacter* and *Escherichia*, all of which are Gram-negative rod bacteria possessing multiple flagella, penetrated the inner mucus layer and were present adjacent to the epithelial cell layers of *Lypd8*^{-/-} mice.

Proteus and *Helicobacter* are associated with the pathogenesis of inflammatory bowel diseases in both mice and humans^{14–16}. Therefore, we analysed the sensitivity of *Lypd8*^{-/-} mice to intestinal

inflammation. Administration of DSS to *Lypd8*^{-/-} mice induced severe intestinal inflammation with high mortality (Fig. 3a–c). Mice were then orally treated with gentamicin, which is active against Gram-negative bacteria including *P. mirabilis* (Extended Data Fig. 6a). The faeces of *Lypd8*^{-/-} mice treated with gentamicin had decreased numbers of *Proteus*, *Escherichia* and *Helicobacter* (Fig. 3d), and, accordingly, no bacteria were present just above the epithelial layers (Fig. 3e, f). Furthermore, gentamicin treatment ameliorated the intestinal inflammation in DSS-treated *Lypd8*^{-/-} mice (Extended Data Fig. 6b, d, f). In contrast, treatment with vancomycin, which is active against Gram-positive cocci but not *P. mirabilis* (Extended Data Fig. 6a), did not induce changes in the inner layers of *Lypd8*^{-/-} mice (Fig. 3e, f). Furthermore, vancomycin treatment resulted in increased numbers of faecal *Lactobacillus*, *Proteus*, *Escherichia* and *Helicobacter* (Fig. 3d) and worsened DSS-induced intestinal inflammation in *Lypd8*^{-/-} mice (Extended Data Fig. 6c, e, g). Thus, mucus penetration by flagellated Gram-negative bacteria such as *Proteus*, *Escherichia* and *Helicobacter* correlates with increased sensitivity to intestinal inflammation in *Lypd8*^{-/-} mice.

We next addressed the mechanisms by which *Lypd8* inhibits bacterial invasion of the epithelial layer. Recombinant *Lypd8* was incubated with faecal samples from wild-type mice (Fig. 4a). *Lypd8* bound to a subpopulation of faecal bacteria. Then, faecal suspensions of wild-type and *Lypd8*^{-/-} mice were incubated with anti-*Lypd8* mAb and analysed by flow cytometry and fluorescence microscopy (Fig. 4b, c). *Lypd8*⁺ bacteria were present in faeces from wild-type mice. Next, we purified *Lypd8*-bound and unbound faecal bacteria by flow cytometry, and determined the bacterial genera present (Fig. 4d). *Escherichia* and *Helicobacter* were substantially enriched in *Lypd8*-bound bacteria, indicating that *Lypd8* preferentially associates with these flagellated bacteria. Flow cytometric analysis further showed that *Lypd8*

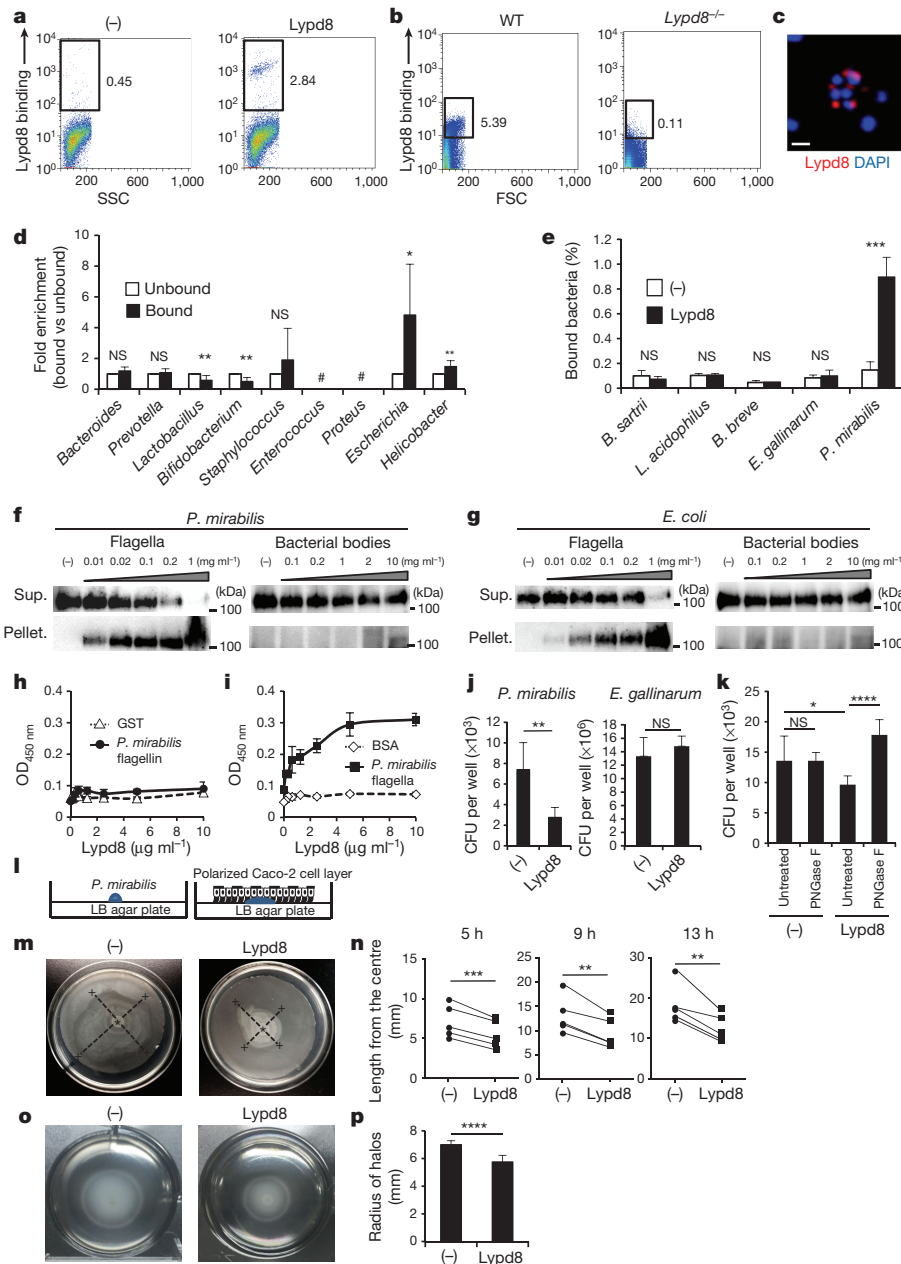


Figure 4 | Lypd8-dependent suppression of *P. mirabilis* attachment to epithelial cells. **a**, Flow cytometry of faecal bacteria incubated with Lypd8. **b**, **c**, Flow cytometry (**b**) and confocal microscopy (**c**) of faecal bacteria incubated with anti-Lypd8 antibody. Faecal bacterium, blue; Lypd8, red. Scale bar, 1 μ m. **d**, Quantitative PCR of Lypd8-bound and unbound bacteria. Data show the fold enrichment of DNA amount of Lypd8-bound versus Lypd8-unbound bacteria ($n = 8$ per group). ND, not detected in unbound fraction. **e**, The percentages of anti-Flag-antibody-bound bacteria incubated with Flag-tagged Lypd8. **f**, **g**, Immunoblot analysis with anti-Flag antibody of the supernatant and pellet of the mixture of Flag-tagged Lypd8 and flagella or bacterial bodies from *P. mirabilis* (**f**) and *E. coli* (**g**). **h**, **i**, ELISA assay of Lypd8 binding to

flagellin (**h**) or flagella (**i**) of *P. mirabilis*. **j**, **k**, The mean CFU of *E. gallinarum* or *P. mirabilis* adhering to Caco-2 cells ($n = 6$) (**j**) and Caco-2 cells pretreated with PNGaseF (**k**). Data are mean \pm s.d. and representative of three independent experiments. **l**–**n**, Schema of swarming motility assay of *P. mirabilis* (**l**). Circular swarms by *P. mirabilis* in the presence of polarized Lypd8-expressing Caco-2 cells at 13 h. *, centre; +, edges of circular swarms. Representative photos (**m**) and the mean length of the circular swarms at the indicated time points ($n = 5$) (**n**). **o**, **p**, Motility of *P. mirabilis* in semisolid agar with or without Lypd8. Representative photos (**o**) and the radii of motility halos at 4 h (**p**). Data are mean \pm s.d. ($n = 6$ per group). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$, **** $P < 0.001$.

bound to pure preparations of *in-vitro*-cultured *P. mirabilis*, but not non-flagellated bacteria such as *Bacteroides sartrii*, *Lactobacillus acidophilus*, *Bifidobacterium breve* and *Enterococcus gallinarum* (Fig. 4e and Extended Data Fig. 7a). However, Lypd8 exhibited no bactericidal activity against *P. mirabilis* (Extended Data Fig. 7b). *P. mirabilis* was incubated with recombinant Lypd8 and analysed by immunogold scanning electron microscopy (Extended Data Fig. 7c). Immunogold-like spheres were observed on flagella of *P. mirabilis* treated with Lypd8.

Therefore, flagella and bacterial bodies were isolated from *P. mirabilis* (Extended Data Fig. 8a, b), incubated with a solution containing recombinant Lypd8, and pelleted down (Fig. 4f). After incubation with increasing flagella concentrations, the amount of recombinant Lypd8 was decreased in the supernatants, but elevated in the pellets. The incubation with bacterial bodies did not alter the amount of Lypd8. Similar observations were made when flagella and bacterial bodies were isolated from commensal *E. coli* (Fig. 4g). When a protein structurally

related to Lypd8 (LOC69864, encoded by *1810065E05Rik*) was used, the protein amounts were not altered by incubation with flagella of *P. mirabilis*, indicating that Lypd8 binds specifically to bacterial flagella (Extended Data Fig. 8c). The main component of flagella is flagellin. Therefore, we performed a quantitative enzyme-linked immunosorbent assay (ELISA) to analyse the association of Lypd8 both with flagellin and flagella (Fig. 4h, i and Extended Data Fig. 8d, e). Unlike TLR5, Lypd8 did not bind to flagellin, but associated with flagella in a dose-dependent manner. As Lypd8 interacts with flagella of live bacteria (Fig. 4a–e, Extended Data Fig. 7a), Lypd8 is suggested to recognize the higher order structure of the flagellum comprised of polymerized flagellin. As Caco-2 cells did not express endogenous *LYPD8* (Extended Data Fig. 9a), we used Caco-2 cells with enforced expression of mouse Lypd8 for further analyses. *P. mirabilis* was added to cultures of polarized Caco-2 cells with or without Lypd8 expression, and after 8 h the bacteria that had attached to Caco-2 cells were analysed (Fig. 4j). In Caco-2 cells expressing Lypd8, attachment of *P. mirabilis*, but not *E. gallinarum*, was markedly diminished. PNGase F treatment of Caco-2 cells expressing Lypd8 increased the number of *P. mirabilis*, indicating the possible involvement of glycan in the inhibitory activity of Lypd8 (Fig. 4k). We added *in-vitro*-cultured *P. mirabilis* to the centre of lysogeny broth agar, then laid the polarized Caco-2 cell sheet with or without Lypd8 over the agar, and incubated for 5, 9 or 13 h (Fig. 4l). Swarming activity of *P. mirabilis* was severely impaired in the presence of Lypd8 (Fig. 4m, n). Because Lypd8 was secreted into culture media by Lypd8-expressing Caco-2 cells (Extended Data Fig. 9b), we investigated whether secreted Lypd8 was able to inhibit bacterial motility. *In-vitro*-cultured *P. mirabilis* was added to the centre of semisolid agar containing recombinant Lypd8, and incubated for 4 h (Fig. 4o, p). In the presence of Lypd8, motility of *P. mirabilis* was reduced. In addition to *P. mirabilis*, commensal *E. coli* migrated into semisolid agar, but its migratory activity was inhibited in the presence of recombinant Lypd8 (Extended Data Fig. 10). These findings indicate that secreted Lypd8 binds to flagella and thereby inhibits bacterial motility in the intestinal lumen.

Our results indicate that Lypd8 promotes the segregation of flagellated bacteria and colonic epithelia, thus reducing the risk of intestinal inflammation. The decreased *LYPD8* expression seen in ulcerative colitis patients indicates that *LYPD8* is involved in the pathogenesis of inflammatory bowel diseases. Whilst TLR5 is known to recognize flagellin and prevent intestinal inflammation by stabilizing microbiota^{17,18}, Lypd8 recognizes and binds with flagella (polymerized flagellins) of live bacteria, reducing their motility and thereby regulating gut homeostasis. The functional association of Lypd8 and TLR5 is an interesting future issue to be addressed. DSS has been shown to make the inner mucus layer permeable to bacteria⁴. In the absence of Lypd8, the number of flagellated commensal bacteria, which invade the epithelia, increased after DSS treatment resulting in the severe inflammation. As enteric pathogenic bacteria harbour flagella, Lypd8 may also have a role in prevention of enteric infection.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 March 2015; accepted 5 February 2016.

Published online 30 March 2016

- McGuckin, M. A., Linden, S. K., Sutton, P. & Florin, T. H. Mucin dynamics and enteric pathogens. *Nature Rev. Microbiol.* **9**, 265–278 (2011).
- Johansson, M. E. *et al.* Composition and functional role of the mucus layers in the intestine. *Cell. Mol. Life Sci.* **68**, 3635–3641 (2011).
- Johansson, M. E. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl Acad. Sci. USA* **105**, 15064–15069 (2008).
- Johansson, M. E., Larsson, J. M. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc. Natl Acad. Sci. USA* **108** (Suppl 1), 4659–4665 (2011).
- Maynard, C. L., Elson, C. O., Hatton, R. D. & Weaver, C. T. Reciprocal interactions of the intestinal microbiota and immune system. *Nature* **489**, 231–241 (2012).
- Peterson, L. W. & Artis, D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nature Rev. Immunol.* **14**, 141–153 (2014).
- Ayabe, T. *et al.* Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria. *Nature Immunol.* **1**, 113–118 (2000).
- Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126–1130 (2006).
- Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L. & Hooper, L. V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc. Natl Acad. Sci. USA* **105**, 20858–20863 (2008).
- Vaishnava, S. *et al.* The antibacterial lectin RegIII γ promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255–258 (2011).
- Mukherjee, S. *et al.* Antibacterial membrane attack by a pore-forming intestinal C-type lectin. *Nature* **505**, 103–107 (2014).
- Zhang, Y. *et al.* Identification and characterization of human *LYPD6*, a new member of the Ly-6 superfamily. *Mol. Biol. Rep.* **37**, 2055–2062 (2010).
- Kong, H. K. & Park, J. H. Characterization and function of human Ly-6/uPAR molecules. *BMB Rep.* **45**, 595–603 (2012).
- Garrett, W. S. *et al.* Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292–300 (2010).
- Mukhopadhyay, I., Hansen, R., El-Omar, E. M. & Hold, G. L. IBD-what role do Proteobacteria play? *Nature Rev. Gastroenterol. Hepatol.* **9**, 219–230 (2012).
- Hansen, R., Thomson, J. M., Fox, J. G., El-Omar, E. M. & Hold, G. L. Could *Helicobacter* organisms cause inflammatory bowel disease? *FEMS Immunol. Med. Microbiol.* **61**, 1–14 (2011).
- Carvalho, F. A. *et al.* Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice. *Cell Host Microbe* **12**, 139–152 (2012).
- Cullender, T. C. *et al.* Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. *Cell Host Microbe* **14**, 571–581 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Fujioka, T. Kondo and Y. Magota for technical assistance, T. Oida for generation of mAb, Y. Matsunaga and J. Takagi for the binding assay, S. Pareek for critical reading of the manuscript, and C. Hidaka for secretarial assistance. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology, the Japan Agency for Medical Research and Development.

Author Contributions R.O. planned and performed experiments. H.K. and M.Y. generated mutant mice. M.K., T.Ki., N.K., T.Kus., Y.U. and E.U. performed animal experiments. T.Kur., D.M., K.G., S.N. and T.I. analysed composition of microbiota. T.N., H.W. and K.S. performed scanning electron microscopy. Y.O. generated *Lypd8*^{venus} mice. H.I., S.B., H.O., H.M. and J.N. performed experiments using human samples. K.T. planned and directed the research and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.T. (ktakeda@ongene.med.osaka-u.ac.jp).

METHODS

Mice. C57BL/6J mice were purchased from Japan SLC (Shizuoka), and ICR mice and germ-free (IQ1/jic[Gf] ICR) mice were from CLEA Japan (Tokyo). *Myd88*^{-/-} and *Asc*^{-/-} mice backcrossed onto C57BL/6 mice for eight or more generations were used^{19–21}. All mice were kept under SPF conditions at the Experimental Animal Facility, Graduate School of Medicine, Osaka University. All animal experiments were performed according to guidelines of the Animal Research Committee of the Graduate School of Medicine at Osaka University.

Quantitative RT-PCR. RNA samples were prepared from indicated tissues using TRIzol reagent (Invitrogen). Total RNA was reverse transcribed using Moloney murine leukaemia virus reverse transcriptase (Promega) and random primers (Toyobo) after treatment with RQ1 DNase I (Promega). cDNA was analysed by real-time RT-PCR using GoTaq qPCR Master Mix (Promega) in ABI 7300 (Applied Biosystems). Values were normalized to the expression of *Gapdh*, and the fold difference in expression relative to that of *Gapdh* is shown. The following primer sets were used: *Lypd8*, 5'-GCCTTCACTGCTCCATCTATTT-3' and 5'-GTGACCATAGCAAGACATGCA-3'; *Villin*, 5'-CTATGCAGATGGTACCTGTTTC-3' and 5'-CCTGGGACGAGTCCTGGCCAA-3'; *Muc2*, 5'-ACATCACCTGTCCCGACTTC-3' and 5'-GAGCAAGGGACTCTGGTCTG-3'; *Lgr5*, 5'-CATCACACTGTCACTGTAGC-3' and 5'-GGTAGCTGACTGATGTTGTTC-3'; *Tnfr*, 5'-TCCAGGCGGTGCCTATGT-3' and 5'-CACCCCGAAGTTCAGTAGACAGA-3'; *Il1b*, 5'-TCAGGACGAGCAGTATCACTCA-3' and 5'-GGAAGGTCCACGGGAAAGAC-3'; *Ifng*, 5'-TCAAGTGGCATAGATGTGGAAGAA-3' and 5'-TGGCTCTGCAGGATTTTCATG-3'; *Il6*, 5'-CTGCAAGAGACTTCCATCCAGTT-3' and 5'-AAGTAGGGAAGGCCGTGGTT-3'; *Cx3d1*, 5'-GGCCGCGTCTTCCATTGT-3' and 5'-TGATAGCGGATGAGCAA-3'; *Cxcl2*, 5'-CTCAGTCTGCAGTGGTCTCTG-3' and 5'-CTGGGGCGCTCACACTCAAGC-3'; *Cd17*, 5'-CTGCAGCATGCCAGAGCT-3' and 5'-GGTCTTATACCAGTCTAC-3'; *Cd28*, 5'-CAGCCTCACCTGAGTCATTC-3' and 5'-CAGTGCAACAGCTGGAGGCCA-3'; *Gapdh*, 5'-CCTCGTCCCGTAGACAAAATG-3' and 5'-TCTCCACTTGGCACTGCAA-3'.

In situ hybridization. Plasmids (pCRII; Invitrogen) containing a cDNA fragment of *Lypd8* were used as templates of RNA probes. Digoxigenin (DIG)-labelled antisense or sense probes were prepared with T7 and SP6 RNA polymerase (Ambion), respectively, using the DIG RNA Labelling Mix (Roche Diagnostics). Colonic tissues of C57BL/6J were fixed with 4% paraformaldehyde (PFA). Serial frozen sections were fixed in 4% PFA for 20 min, incubated in cold 0.1% H₂O₂, and permeabilized with 50 µg ml⁻¹ proteinase K for 5 min. After an additional fixation with PFA, the sections were treated with acetic anhydride in triethanolamine for 5 min. The sections were then pre-hybridized with 50% formamide, 5 × saline sodium citrate, 1 mg ml⁻¹ yeast tRNA (Roche Diagnostics), 100 µg ml⁻¹ heparin, 1 × Denhardt's solution, 0.1% Tween 20 at 60 °C for 3 h, and hybridized at 60 °C for 16 h. After washing, sections were incubated with horseradish peroxidase (HRP)-anti-DIG Fab fragment (clone 1.71.256; Roche Diagnostics), followed by biotin-labelled tyramide (TSA Biotin System; Perkin Elmer) for signal amplification. Hybridized probes were detected by ABC-Alkaline phosphatase (Vector Laboratories) and NBT/BCIP (Roche Diagnostics). Images were obtained using BZ-9000 (Keyence).

Generation of *Lypd8*^{venus} mice and *Lypd8*^{-/-} mice. Targeting vectors were constructed by replacement of genomic fragment containing the third and fourth exons of *Lypd8* with a Cre recombinase internal ribosome entry site Venus neomycin-resistance gene cassette (*Lypd8*^{venus} mice) or a neomycin-resistance gene cassette (*Lypd8*^{-/-} mice), and a gene encoding HSV thymidine kinase driven by a phosphoglycerate kinase promoter inserted into the genomic fragment. After the targeting vector was transfected into V6.5 embryonic stem cells, G418 and ganciclovir double-resistant colonies were selected and screened by PCR and Southern blot analysis. Homologous recombinants were used for generation of *Lypd8*^{venus} mice and *Lypd8*^{-/-} mice. *Lypd8*^{venus} mice and *Lypd8*^{-/-} mice were backcrossed onto C57BL/6 mice for at least six generations, and *Lypd8*^{venus} mice, *Lypd8*^{-/-} mice and their wild-type littermates from intercrosses of heterozygous mice were kept in the same cages and used for experiments. There was no randomization, but stratification was used to achieve the similar ages and sex ratios among experimental groups. The experiments were not blinded. Flow cytometry of isolated colonic epithelial cells

Intestinal epithelial cells were isolated from *Lypd8*^{venus} mice by shaking intestinal tissues in 5 mM EDTA/HBSS solution at 37 °C for 20 min and then washed with phosphate buffered saline (PBS). The intestinal epithelial cells were treated with PE-Cy7-conjugated anti-CD3e Ab (clone 145-2C11; BD Biosciences) in PBS containing 2% FBS to block nonspecific binding. Flow cytometric analysis was performed using a FACSCanto II flow cytometer (BD Biosciences) with FlowJo software (Tree Star). For cell isolation, cells were sorted using a FACSaria (BD Biosciences).

Generation of colonic epithelial cell lines stably expressing *Lypd8*. A sequence for Flag-tagged *Lypd8* was constructed by inserting Flag-tag sequence into total *Lypd8* coding sequence immediately downstream of predicted N-terminal signal sequence. CMT-93 cells and Caco-2 cells, which were originally obtained from ATCC and free of mycoplasma, were transfected with linearized pcDNA3.1 (+) vector (Invitrogen) inserted the sequence for Flag-tagged *Lypd8* using Lipofectamine2000 (Invitrogen). These cells were cultured in G418-containing medium. The surviving cells were stained with anti-Flag M2 monoclonal antibody (cat F3165; Sigma-Aldrich) and Alexa Fluor 488 goat anti-mouse IgG antibody (cat A11001; Molecular Probes) and cells expressing Flag-tagged *Lypd8* were sorted using FACSaria (BD Biosciences).

Immunostaining of Caco-2 cell line cultured on transwell membrane. Caco-2 cells stably expressing Flag-tagged *Lypd8* (1 × 10⁵ cells) were cultured on transwell filters of 3.0-µm pore size (BD Biosciences) for 2 weeks. After confirming full confluency, Caco-2 cells were washed with PBS, fixed with 4% PFA and then blocked by 1% bovine serum albumin (BSA) in PBS. Cells were stained with mouse anti-Flag M2 mAb (Sigma-Aldrich) plus Alexa Fluor 488 goat anti-mouse IgG (Invitrogen) or rabbit anti-human Claudin-1 Ab (cat ab15098; Abcam) plus Alexa Fluor 594 goat anti-rabbit IgG (cat A11012; Invitrogen), and counterstained with DAPI (Vector Laboratories). Transwell filters with these cells were cut, placed on slide glasses and analysed using a confocal microscope (FV1000-D; Olympus). The supernatants on Caco-2 cells with or without *Lypd8* expression were incubated with anti-Flag M2 affinity gel (Sigma-Aldrich) at 4 °C for 3 h. The resin was collected, washed three times with Tris-buffered saline and then suspended in sample buffer for immunoblot analysis using anti-Flag M2 mAb (Sigma-Aldrich) and HRP-conjugated goat anti-mouse IgG (cat NA931; GE Healthcare).

Assay for GPI cleaving activity. CMT93 cells stably expressing *Lypd8* (1 × 10⁶ cells) were rinsed twice with cold PBS and incubated with 0.5 ml of the same buffer containing 0.5 units of *Bacillus cereus* phosphatidylinositol-specific phospholipase C (Invitrogen) at 4 °C for 20 min. These cells were stained with anti-Flag M2 mAb (Sigma-Aldrich) and Alexa Fluor 488 goat anti-mouse IgG (Invitrogen). The surface expression of *Lypd8* was analysed using a FACSCanto II (BD Biosciences).

Generation of anti-*Lypd8* antibody. *Lypd8*^{-/-} mice were immunized with P3U1 cells with retroviral overexpression of *Lypd8*. Lymph node cells from immunized *Lypd8*^{-/-} mice were fused with Sp2/0 mouse myeloma cells. Hybridomas were screened by analysing reactivity to RBL1 cells overexpressing Flag-tagged *Lypd8* with flow cytometry. Positive clones were selected for further subcloning. Culture supernatants of subcloned hybridoma cells were screened by immunostaining of 4% PFA- or Carnoy's-fixed colon sections from wild-type and *Lypd8*^{-/-} mice. Culture supernatants of hybridoma clones (clone number 4F8 and 5A4), which stained Carnoy's-fixed colon sections of wild-type mice, but not *Lypd8*^{-/-} mice, were used as mouse anti-mouse *Lypd8* mAb. The specificity of mAb was also confirmed by flow cytometry analysis of CMT93 cells with or without *Lypd8* expression. Anti-mouse *Lypd8* mAb was purified from the supernatant by Ex-pure Spin ProG (Kyoto monotech) and labelled with CF633 using Mix-n-Stain CF633 antibody labelling kit (Biotium).

Immunostaining and fluorescence in situ hybridization. Colons from 8- to 12-week-old mice without washing were fixed in methanol-Carnoy's fixative composed of 60% methanol, 30% chloroform and 10% acetic acid. Paraffin-embedded sections were dewaxed and hydrated. Sections were blocked with 1% BSA in PBS and stained with CF633-conjugated anti-mouse *Lypd8* mAb or anti-Mucin2 Ab (clone H-300; Santa Cruz Biotechnology) and Alexa Fluor 594 goat anti-rabbit IgG (cat A11012; Invitrogen). Sections were incubated with 1 µg Cy3- or Cy5-conjugated EUB338 (5'-GCTGCCTCCCGTAGGAGT-3') for detection of all bacteria or Cy3-conjugated pB-02110 (5'-ATGGGTTTCATCCCATAGTGC-3')¹⁴ for detection of *Proteus* spp. in 200 µl of hybridization buffer (750 mM NaCl, 100 mM Tris-HCl (pH 7.4), 5 mM EDTA, 0.01% BSA, 10% dextran sulfate) at 40 °C for 16 h. Sections were rinsed in wash buffer (50 mM NaCl, 4 mM Tris-HCl (pH 7.4), 0.02 mM EDTA), washed at 45 °C for 20 min and counterstained with DAPI (Vector Laboratories). Human colons from the normal mucosa of resected colon tissues from colon cancer patients with no history of a diagnosis of inflammatory bowel diseases or the inflamed mucosa of resected colon tissues from ulcerative colitis patients were fixed in 4% PFA. The experiment was approved by the Ethical Committee of Osaka University School of Medicine. All participants provided informed consent. All ulcerative patients had a confirmed diagnosis by a gastroenterologist. Paraffin-embedded sections were dewaxed and hydrated. Sections were blocked with 1% BSA in PBS and stained with anti-human *Lypd8* Ab (clone V-16; Santa Cruz Biotechnology) and Alexa Fluor 568 donkey anti-goat IgG (cat A11057; Invitrogen). Normal goat IgG (cat AB-108-C; R&D systems) was used as an isotype control. Sections were analysed using a confocal microscope (FV1000-D; Olympus). The distance between bacterial populations and the epithelial surface was measured at four points of the proximal, middle and distal colon in each mouse.

Purification of Flag-tagged recombinant proteins. Recombinant Lypd8 or LOC69864 (an uncharacterized protein that is structurally most similar to Lypd8) was purified from CMT93 cells stably expressing FLAG-tagged Lypd8 or LOC69864 using FLAG M Purification Kit (Sigma-Aldrich). As a negative control, cells transfected with empty vector (pcDNA3.1 (+) (Invitrogen)) were used.

Deglycosylation assay. Recombinant Lypd8 proteins (1 µg) were incubated with PNGase F, sialidase A and O-glycanase (Prozyme) at 37 °C for 3 h. Recombinant Lypd8 proteins treated with glycanase were separated with SDS-PAGE and transferred to polyvinylidene fluoride membranes (Millipore) that were incubated with anti-Flag M2 mAb (Sigma-Aldrich) and then HRP-conjugated goat anti-mouse IgG (cat NA931: GE Healthcare). Immunoreactivity was detected using SuperSignal (Thermo Scientific).

Analysis of mutant Lypd8 (N-D) protein. The mutant Lypd8 N-D sequence was designed so that thirteen Asp (N) residues were converted to Asn (D). 293T cells were transfected with the mutant Lypd8 (N-D) expression vector using Lipofectamine 2000 (Invitrogen) and cell lysates of 293T cells expressing mutant Lypd8 (N-D) protein were separated with SDS-PAGE and analysed by western blot.

Extraction of bacterial DNA. Faeces, luminal contents of the colon or colonic tissues were collected in tubes containing RNAlater (Ambion). After weights were measured, RNAlater was added to make tenfold dilutions of homogenates. Homogenates (200 µl) of faeces, luminal contents or colonic tissues, or bacterial suspension containing sorted bacteria were washed twice with 1 ml PBS, 0.3 g glass beads (diameter, 0.1 mm) (BioSpec Products), 300 µl Tris-SDS solution and 500 µl TE-saturated phenol were added to the suspension, and the mixture was vortexed vigorously using a FastPrep-24 (M.P. Biomedicals) at 5.0 power level for 30 s. After centrifugation at 20,000 g for 5 min at 4 °C, 400 µl of supernatants were collected. Subsequently, phenol-chloroform extraction was performed and 250 µl of supernatants were subjected to isopropanol precipitation. Finally, DNAs were suspended in 200 µl TE buffer and stored at -20 °C.

Determination of microbiota by deep sequencing. PCR was performed using a primer set (784F, 5'-AGGATTAGATACCTGGTA-3'; and 1061R, 5'-CRRACGAGCTGACGAC-3') targeting the V5-V6 region of the 16S rRNA genes with KAPA HiFi HotStart Ready Mix (KAPA Biosystems). Products were purified using DNA clean and Concentrator-5 (Zymo Research). Adaptor and barcode sequences were attached to the products by 10 cycles of PCR with 1 ng of each of initial PCR product as the template and primer sets and other PCR conditions were unchanged. Sequencing was performed using a 316 chip and Ion PGM Sequencing 400 Kit (Life Technologies) on the Ion PGM sequencer (Life Technologies). Raw sequences were demultiplexed and quality-trimmed by the following procedures: (1) trimming bases with quality below Q15 from 3' end of each read, (2) removing reads with average quality below Q20, (3) removing reads without primer sequences on both ends, and (4) removing reads with length shorter than 260 basepairs, using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and BBtrim (<http://bbmap.sourceforge.net/>). The processed sequences were then clustered into operational taxonomic units (OTU) defined at 94% similarity cutoff using UCLUST version 1.2.22q. Representative sequences for each OTU were classified taxonomically using RDP Classifier version 2.2 with the SILVA 111 database.

Quantitative PCR amplification of 16S rRNA genes. Quantitative PCR was performed in ABI7300 using GoTaq qPCR Master Mix (Promega). The following primer sets were used: 'all bacteria', 5'-CGG TGAATACGTTCCCGG-3' and 5'-TACGGCTACCTTGTACGACTT-3'; *Bacteroides*, 5'-GAGAGGAAGGTCCCCAC-3' and 5'-CGCTACT TGGCTGGTTCAG-3'; *Prevotella*, 5'-CACRGTAACGATGGATGCC-3' and 5'-GGTCGGGTTGCAGACC-3'; *Lactobacillus*, 5'-TGGAACA GRTGCTAATACCG-3' and 5'-GTCCATTGTGGAAGATTCCC-3'; *Bifidobacterium*, 5'-CTCCTGGAAACGGGTGG-3' and 5'-GGTGTTCTT CCCGATATCTACA-3'; *Escherichia/Shigella*, 5'-GAGTAAAGTTAAT ACCTTTGCTCATTG-3' and 5'-GAGACTCAAGCTKRCCAGTATCAG-3'; *Helicobacter*, 5'-CTATGACGGGTATCCGGCC-3' and 5'-TCGCCTTCG CAATGAGTATT-3'; *Staphylococcus*, 5'-TTTGGGCTACACGCTGCTAC AATGGACAA-3' and 5'-AACAACTTTATGGGATTTGCWTGA-3'; *Enterococcus*, 5'-ATCAGAGGGGATAACACTT-3' and 5'-ACTCTCATCCT TGTCTTCTC-3'; and *Proteus*, 5'-GTTATTCGTGATGGTATGGG-3' and 5'-ATAAAGGTGGTTACGCCAGA-3'. In the analysis of Lypd8-bound and unbound bacteria, all samples were normalized to the 16S rRNA gene level of 'all bacteria'.

Culture of colonic tissue homogenates. Colonic tissues were thoroughly washed twice and collected in tubes containing 5 ml PBS. After colonic tissues were homogenized in PBS using a homogenizer, 100 µl of the homogenates were incubated on lysogeny broth (LB) or MacConkey agar plates at 37 °C in Ruskinn Bugbox Plus

anaerobic chamber (The Baker Company) for 24 h. The number of swarming colonies on LB agar plates or white colonies on MacConkey agar plates was counted at 16 h after the start of incubation and colony-forming units were calculated.

Inoculation of CFSE-labelled *P. mirabilis*. *P. mirabilis* was cultured in LB medium at 37 °C until OD₆₀₀ reached 0.6. The bacterial culture was centrifuged at 8,000g for 5 min and the pellet was resuspended in 1 ml PBS. The bacterial suspension was re-centrifuged and the pellet was labelled with CFSE Fluorescent Cell Labelling Kit (Abcam). After confirming CFSE labelling of *P. mirabilis* using a FACSCanto II (BD Biosciences), CFSE-labelled *P. mirabilis* was suspended with PBS (5 × 10⁸ per ml) and the bacterial suspension (200 µl) was inoculated into anaesthetized mice via the transanal route with a catheter. At 1, 2, 4, and 8 h after inoculation, colons without washing were fixed in 4% PFA. Colon sections were counterstained with DAPI (Vector Laboratories), and analysed using a confocal microscope (FV1000-D; Olympus).

Induction of DSS-induced colitis. Eight- to ten-week-old *Lypd8*^{-/-} mice and their littermate wild-type mice were used for DSS-induced colitis experiments. Acute colitis was induced by administration of 2% DSS (36–50 kDa; MP Biomedicals) in the drinking water for 5 days. Mice were analysed for changes in weight, survival rates and histological changes.

Antibiotic treatment. Mice were treated with gentamicin (2 g l⁻¹; Nacalai Tesque) or vancomycin (500 mg l⁻¹; Duchefa Biochemie B.V.) dissolved in autoclaved drinking water and provided for 2 weeks. Fluid intake was monitored.

Assay of Lypd8 binding to faecal bacteria. Faeces were collected in tubes containing PBS. After weights were measured, PBS was added to make tenfold diluted suspensions. PBS-diluted faeces were mixed well and centrifuged at 400g for 5 min to remove larger particles from bacteria. Supernatants (200 µl) were centrifuged at 8,000g for 10 min to remove non-bound immunoglobulins. Pellets were resuspended in 1 ml PBS containing 2% FBS, and used as a bacterial suspension. Bacterial suspensions from wild-type mice and *Lypd8*^{-/-} mice were incubated with anti-Lypd8 antibody. After washing with PBS containing 2% FBS, bacterial pellets were stained with Alexa Fluor 647 goat anti-mouse IgG1 (Invitrogen). Bacterial suspensions from C57BL/6J mice were incubated with recombinant Lypd8 protein (1 µg) on ice for 30 min and washed with PBS. Next, bacterial pellets were treated with anti-Flag M2 mAb (Sigma-Aldrich). Finally, bacterial pellets were stained with Alexa Fluor 488 goat anti-mouse IgG (Invitrogen). Bacterial suspensions were analysed using a FACSCanto II (BD Biosciences) with FlowJo software (Tree Star) or a confocal microscope (FV1000-D; Olympus). Lypd8-bound and unbound bacteria were sorted using a FACS Aria (BD Biosciences).

Assay of Lypd8 binding to pure preparations of bacteria. Bacterial strains, *Bacteroides sartrii* JCM 17136^T, *Lactobacillus acidophilus* JCM 1132^T, *Bifidobacterium breve* JCM 1192^T and *Enterococcus gallinarum* JCM 8728^T, were obtained from Japan Collection of Microorganisms. *P. mirabilis* was isolated from the colonic tissue of *Lypd8*^{-/-} mice. All bacteria were cultured in Gifu anaerobic medium (Nissui) under anaerobic conditions. The pure preparation of each bacteria was centrifuged at 8,000g for 5 min and the pellet was resuspended in 1 ml PBS containing 2% FBS. Bacteria were incubated with recombinant Lypd8 protein (1 µg) on ice for 1 h and washed with PBS. Next, the bacterial pellet was treated with anti-Flag M2 mAb (Sigma-Aldrich). Finally, the bacterial pellet was stained with Alexa Fluor 488 goat anti-mouse IgG (Invitrogen) and analysed using a FACSCanto II (BD Biosciences).

Scanning electron microscopy with immunolabelling. Bacterial suspension was mixed with FLAG-tagged recombinant Lypd8 (rLypd8; 20 ng µl⁻¹), and then incubated for 1 h at 4 °C. After the incubation, bacterial suspensions were centrifuged, and pellet was resuspended in PBS. A drop of bacterial suspension was placed on an 3-aminopropyltriethoxysilane (APS)-coated cover slip (Matsunami Glass IND), and blocked with bovine serum albumin and normal goat serum. The mounted bacteria were reacted with anti-FLAG M2 mAb (Sigma-Aldrich), washed with PBS, and reacted with 5-nm-gold-labelled anti-mouse IgG goat antibody (EY Laboratories). The sample was fixed with 2% glutaraldehyde and 1% OsO₄, dehydrated in a graded ethanol series and 3-methylbutyl acetate, and dried in a critical-point drying chamber (HCP-1, Hitachi High-Technologies). It was then coated with a platinum layer approximately 1 nm thick in an ion sputter coater (E-1030, Hitachi High-Technologies), and examined under a scanning electron microscope (S-5000, Hitachi High-Technologies).

Assay of Lypd8 binding to flagella. Bacterial suspension of *P. mirabilis* or *E. coli* JCM 1649^T (Japan Collection of Microorganisms) in PBS was shaken 300 times per min for 60 min to remove flagella from bacterial bodies. Bacterial bodies were pelleted by centrifuging at 4,000g for 20 min. The supernatants were ultracentrifuged at 80,000g for 60 min to obtain flagella. Bacterial bodies or flagella were mixed with the solution of FLAG-tagged recombinant Lypd8 (10 ng µl⁻¹) or LOC69864 (10 ng µl⁻¹), and incubated for 3 h at 4 °C. After the incubation, the bacterial bodies or flagella were pelleted by centrifugation or ultracentrifugation, respectively.

Then, the supernatant was collected and the pellet was resuspended in PBS. The supernatant and pellet suspension were separated with SDS-PAGE and transferred to polyvinylidene fluoride membranes (Millipore) that were incubated with anti-Flag M2 mAb (Sigma-Aldrich) and then HRP-conjugated goat anti-mouse IgG (cat NA931: GE Healthcare). Immunoreactivity was detected using SuperSignal (Thermo Scientific).

Purification of recombinant flagellin protein of *P. mirabilis*. Bacterial DNA of *P. mirabilis* was extracted from the culture of *P. mirabilis* using a previously described protocol (see above). Using the extracted bacterial DNA as template, PCR was performed to isolate the flagellin gene with a primer set (5'-GGATCCGCACAAAGTTATTAATACTAATTAT-3' and 5'-GCGGCCGCTTAACGTAACAGACAGAACAGT-3') targeting the full-length flagellin gene of *P. mirabilis*. The isolated DNA was cloned in frame into EcoRI- and NotI-digested pGEX-6P-2 (GE Healthcare) to generate a glutathione S-transferase (GST)-flagellin fusion construct. A recombinant flagellin protein was prepared by growing *E. coli* harbouring the pGEX-flagellin plasmids in LB medium containing 100 µg ml⁻¹ of ampicillin at 37 °C until OD₆₀₀ reached 0.6. After growth, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the medium, and bacterial culture was continued for 4 h at 37 °C. Cells were centrifuged and the cell pellet was suspended in PBS with 1% Triton X-100. The cell suspension was sonicated 5 times for 30 s at 4 °C, and cleared by centrifugation. The resulting solution was bound to Glutathione-Sepharose 4 Fast Flow (GE Healthcare). The beads were washed 3 times. The GST fusion proteins were then eluted with elution buffer (5 mM glutathione, 50 mM Tris-HCl, pH 9.6). Elution buffer was replaced with PBS with Amicon Ultra-15 (Millipore). The amounts and purity of the protein were estimated by SDS-PAGE.

ELISA binding assay. 96-well plates (Corning) were coated with 5 µg ml⁻¹ GST peptides, 5 µg ml⁻¹ GST-tagged flagellin protein of *P. mirabilis*, 100 µg ml⁻¹ flagella of *P. mirabilis* or PBS 1% BSA for 16 h at 4 °C. Plates were then washed, and increasing concentrations of mouse TLR5 Fc protein (R&D systems) or FLAG-tagged Lypd8 protein diluted in PBS 1% BSA were added and incubated for 2 h at room temperature. Plates were then washed, and 0.5 µg ml⁻¹ of HRP-conjugated goat anti-mouse IgG (cat NA931: GE Healthcare) or HRP-conjugated anti-FLAG M2 mAb (Sigma-Aldrich) diluted in PBS 1% BSA was added and incubated for 2 h at room temperature. Plates were then washed, and TMB substrate was added. Plates were then read at 450 nm with a spectrometer.

In vitro growth assay of *P. mirabilis*. *P. mirabilis* (1 × 10⁵) were incubated in LB medium (1 ml) with the indicated concentrations of recombinant Lypd8 protein for 6 h at 37 °C. The bacterial cultures were applied to MacConkey agar plates and incubated at 37 °C for 16 h, and the number of colonies was counted and colony-forming units were calculated.

Analysis of LYPD8 expression in Caco-2 cells and human colon epithelia. Human colon epithelia were obtained from the normal mucosa of resected colon tissues from colon cancer patients. Colonic epithelial layers were first stripped by shaking sections of colon in 5 mM EDTA/HBSS. Total RNA was isolated from Caco-2 cells and human colon epithelia and reverse transcribed. cDNA was analysed by real-time RT-PCR. Values were normalized to the expression of *GAPDH*, and the fold difference in expression relative to that of *GAPDH* is shown. The following primer sets were used: *LYPD8*, 5'-GAACACTTTCATTTTGTAAGC-3' and 5'-ACGACAGGAAGTTCCATTAGA-3'; *GAPDH*, 5'-TGGATATTGTTGCCATCAATG-3' and 5'-TGATGGGATTTCATTGA

TGA-3'. The experiment was approved by the Ethical Committee of Osaka University School of Medicine and informed consent for specimen use was obtained from all patients.

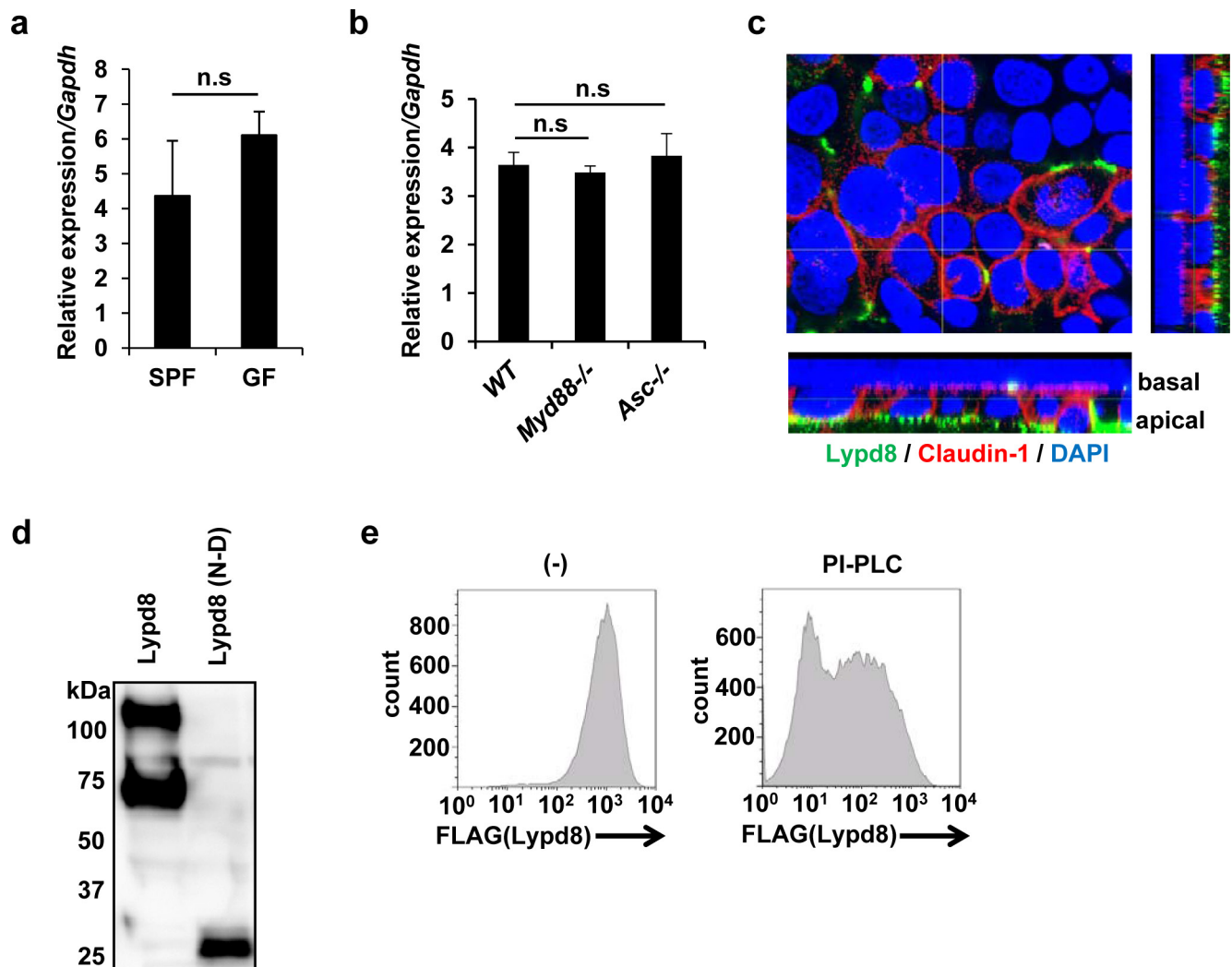
Bacterial adhesion assay on Caco-2 cells. *E. gallinarum* or *P. mirabilis* (1 × 10⁵) was added to a culture of polarized Caco-2 cells. For PNGase F treatment, cells were irradiated with 30 Gy to stop cell growth and then incubated in culture medium containing PNGaseF (Sigma-Aldrich) (2 unit/ml) for 2 h. At 8 h after the addition of *P. mirabilis*, each well was washed with PBS three times and then Caco-2 cells were removed from transwell filters with trypsin/EDTA solutions and suspended in PBS. PBS containing Caco-2 cells were applied to LB or MacConkey agar plates and incubated at 37 °C for 16 h, and the number of colonies was counted and colony-forming units were calculated.

Swarming motility assay of *P. mirabilis*. *P. mirabilis* was cultured in LB medium at 37 °C until OD₆₀₀ was 0.6. LB agar plates (1.5%) were centrally inoculated with 5 µl of bacterial culture and incubated at 37 °C. After 1 h, polarized Caco-2 cells with or without Flag-tagged Lypd8 expression, cultured on 75 mm diameter transwell filters of 3.0 µm pore size (BD Biosciences) for approximately 2 weeks, were added to *P. mirabilis*-inoculated agar plate so that the apical side of cells adhered to the surface of the agar plate. Motility was assessed by examining circular swarms formed by growing motile bacterial cells. The length from the centre to four points of the edge of circular swarms was measured at 5, 9 and 13 h time points following *P. mirabilis* inoculation.

Motility assay of *P. mirabilis* and *E. coli* in semisolid agar. CMT-93 cells (5 × 10⁷ cells) with or without Flag-tagged Lypd8 were lysed with 1 ml of CellLytic M Cell Lysis Reagent (Sigma-Aldrich). After removal of any insoluble material by centrifugation and filtration with a 0.45 µm filter, the cell lysates were incubated with 50 µl of anti-Flag M2 affinity gel (Sigma-Aldrich) for 3 h. The resin was centrifuged and washed with Tris-buffered saline three times. The resin containing Flag-tagged Lypd8 was mixed with 1 ml of LB medium containing 0.3% agar. The final concentration of Lypd8 in LB agar was estimated by SDS-PAGE and was approximately 1.25 µg ml⁻¹. *P. mirabilis* and *E. coli* were cultured in LB medium at 37 °C until OD₆₀₀ was 0.6. Semisolid LB agar (0.3%) containing Flag-tagged Lypd8 were centrally inoculated with 0.5 µl of bacterial culture and incubated at 37 °C. Motility was assessed by examining circular migration. The radius of circles formed by bacterial migration was measured at 4 h after the bacterial inoculation.

Statistical analysis. Data are presented as mean ± s.d., as indicated in the figure legends. Differences between control and experimental groups were evaluated using a two-tailed unpaired Student's *t*-test. For the swarming motility assay, a paired Student's *t*-test was performed. A *P* value of <0.05 was considered significant. For the determination of microbiota by deep sequencing, principal component analysis was used to visualize data sets by statistical programming language R 2.1.5. No statistical methods were used to predetermine sample size. No animal or sample was excluded from the analysis.

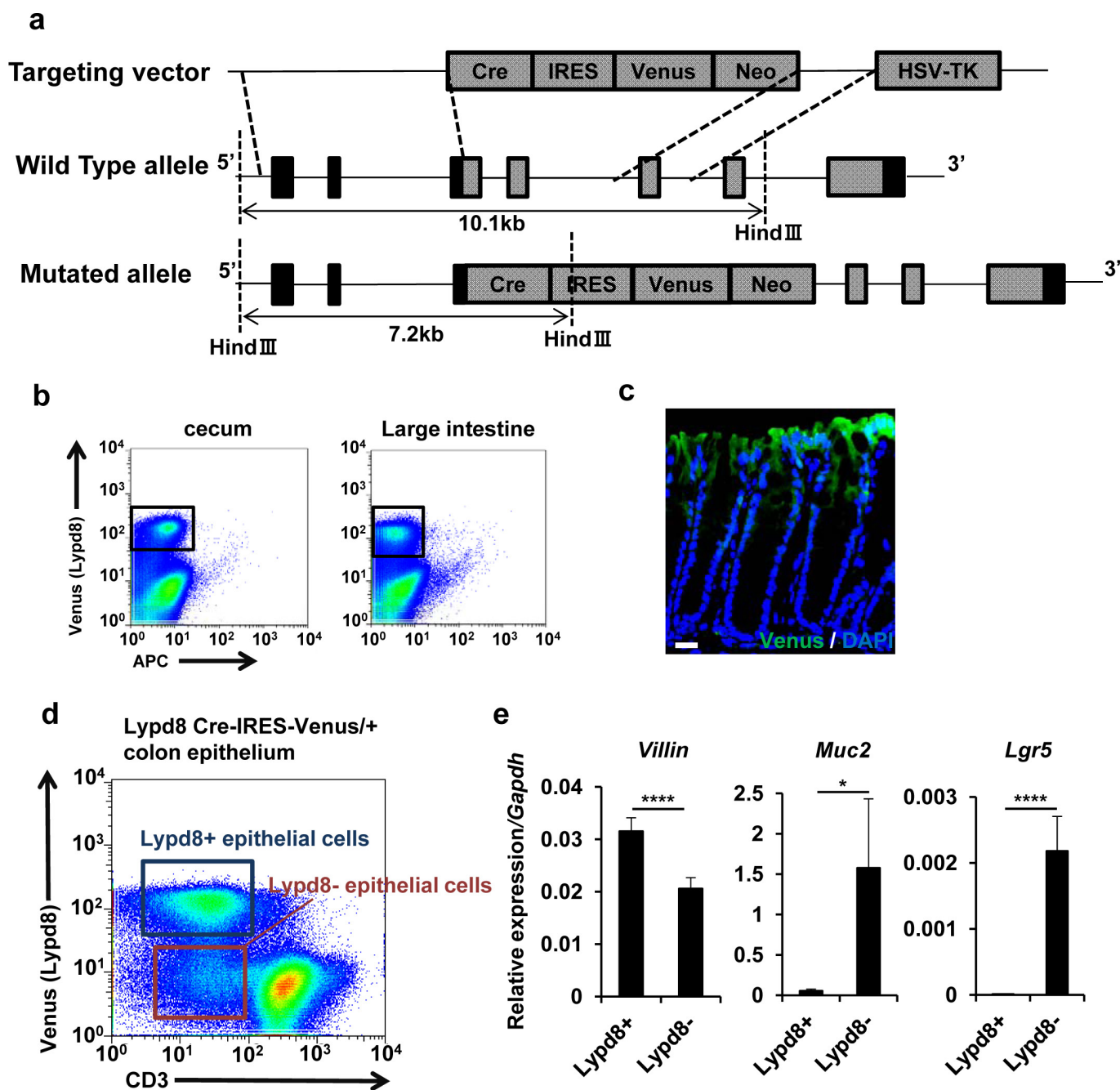
- Adachi, O. *et al.* Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity* **9**, 143–150 (1998).
- Yamamoto, M. *et al.* ASC is essential for LPS-induced activation of procaspase-1 independently of TLR-associated signal adaptor molecules. *Genes Cells* **9**, 1055–1067 (2004).
- Takeda, K. *et al.* Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils. *Immunity* **10**, 39–49 (1999).



Extended Data Figure 1 | Lypd8 is a GPI-anchored glycoprotein.

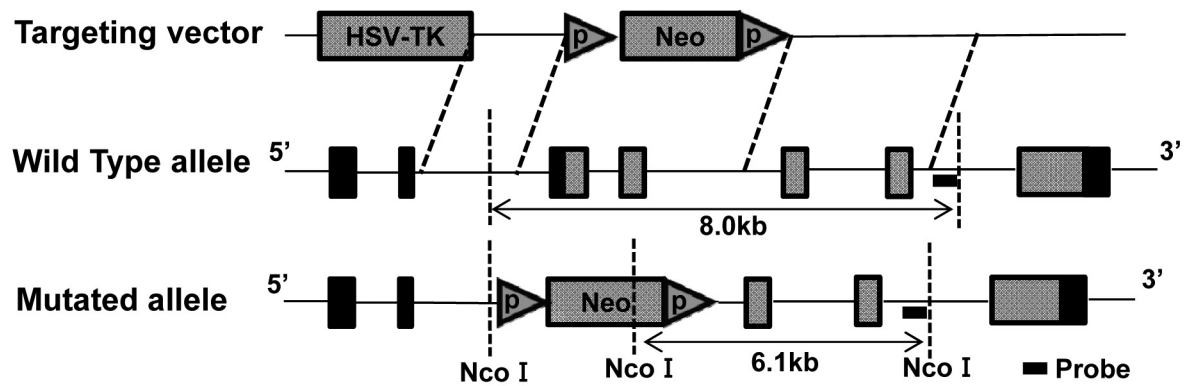
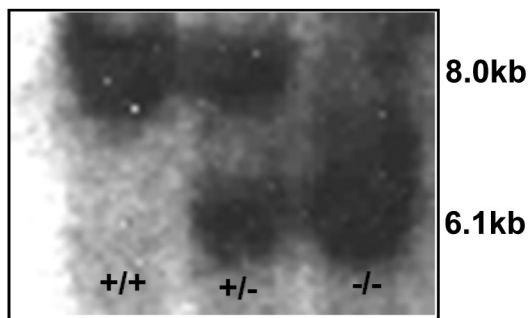
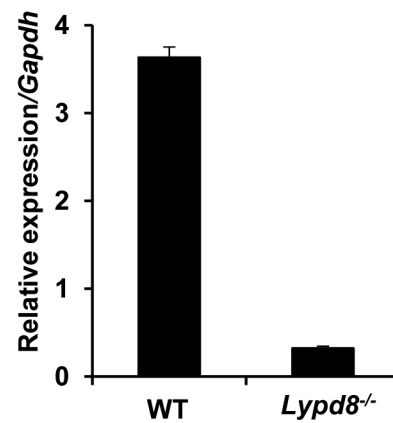
a, b, Quantitative RT-PCR analysis of *Lypd8* expression in the colonic epithelia of ICR mice in specific-pathogen-free and germ-free conditions (**a**), and in colonic epithelia of wild-type, *Myd88*^{-/-} and *Asc*^{-/-} mice (**b**). Values were normalized to *Gapdh*. Data represent the mean \pm s.d. ($n = 3$ per group). NS, not significant. **c**, Confocal microscopic image of Caco-2 cells expressing Flag-tagged Lypd8 cultured on transwell membrane,

stained with anti-Claudin-1 antibody (red), anti-Flag antibody (green), and DAPI (blue). **d**, 293T cell lysates transiently expressing Flag-tagged Lypd8 and Flag-tagged mutant Lypd8 (N-D), in which thirteen Asp residues were converted to Asn, were separated with SDS-PAGE and immunoblotted with anti-Flag antibody. **e**, Flow cytometric analysis of CMT93 cells expressing Flag-tagged Lypd8 before (left) and after (right) treatment with PI-PLC.



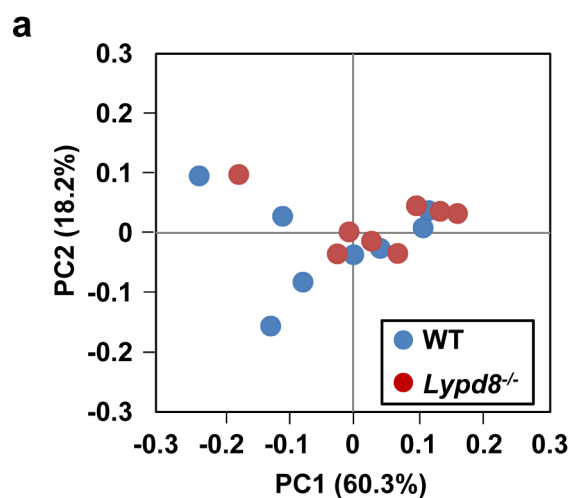
Extended Data Figure 2 | Lypd8 is expressed in enterocytes. **a**, Structures of the *Lypd8* gene (middle), the targeting vector (top) and the predicted Cre-IRES-Venus inserted gene (bottom). Grey box, coding exon; black box, non-coding exon. **b**, Flow cytometry of intestinal epithelial cells isolated from caecum and colon of *Lypd8*^{venus} mice. Venus⁺ epithelial cells (Lypd8⁺) are boxed. **c**, Colon sections of *Lypd8*^{venus} mice were counterstained with DAPI to visualize nuclei. Scale bar, 20 μ m. Venus⁺ Lypd8 was expressed at the uppermost layer of the colon epithelia. **d**, Flow cytometry of intestinal epithelial cells isolated from the colon of *Lypd8*^{venus}

mice stained for CD3. **e**, Lypd8-expressing colonic epithelial cells (black box in **d**, Lypd8⁺) and Lypd8 non-expressing colonic epithelial cells (red box in **d**, Lypd8⁻) are isolated and analysed for expression of *Villin* (highly expressed in enterocytes), *Muc2* (selectively expressed in goblet cells) and *Lgr5* (selectively expressed in intestinal stem cells) by quantitative RT-PCR. Lypd8⁺ cells expressed *Villin*, but neither *Muc2* nor *Lgr5*. Values were normalized to *Gapdh* expression. Data are mean \pm s.d. ($n = 4$ per group). * $P < 0.05$, **** $P < 0.0001$.

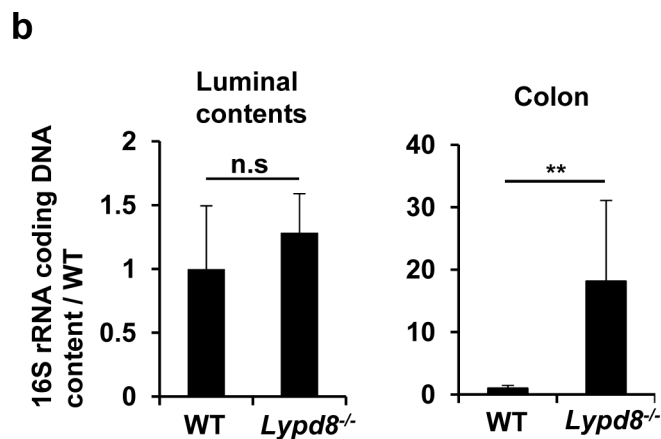
a**b****c**

Extended Data Figure 3 | Generation of *Lypd8*^{-/-} mice. **a**, Structures of the *Lypd8* gene (middle), the targeting vector (top) and the predicted disrupted gene (bottom). Grey box, coding exon; black box, non-coding exon. **b**, Southern blot analysis of offspring from the heterozygote intercrosses. Genomic DNA was extracted from mouse tails, digested with

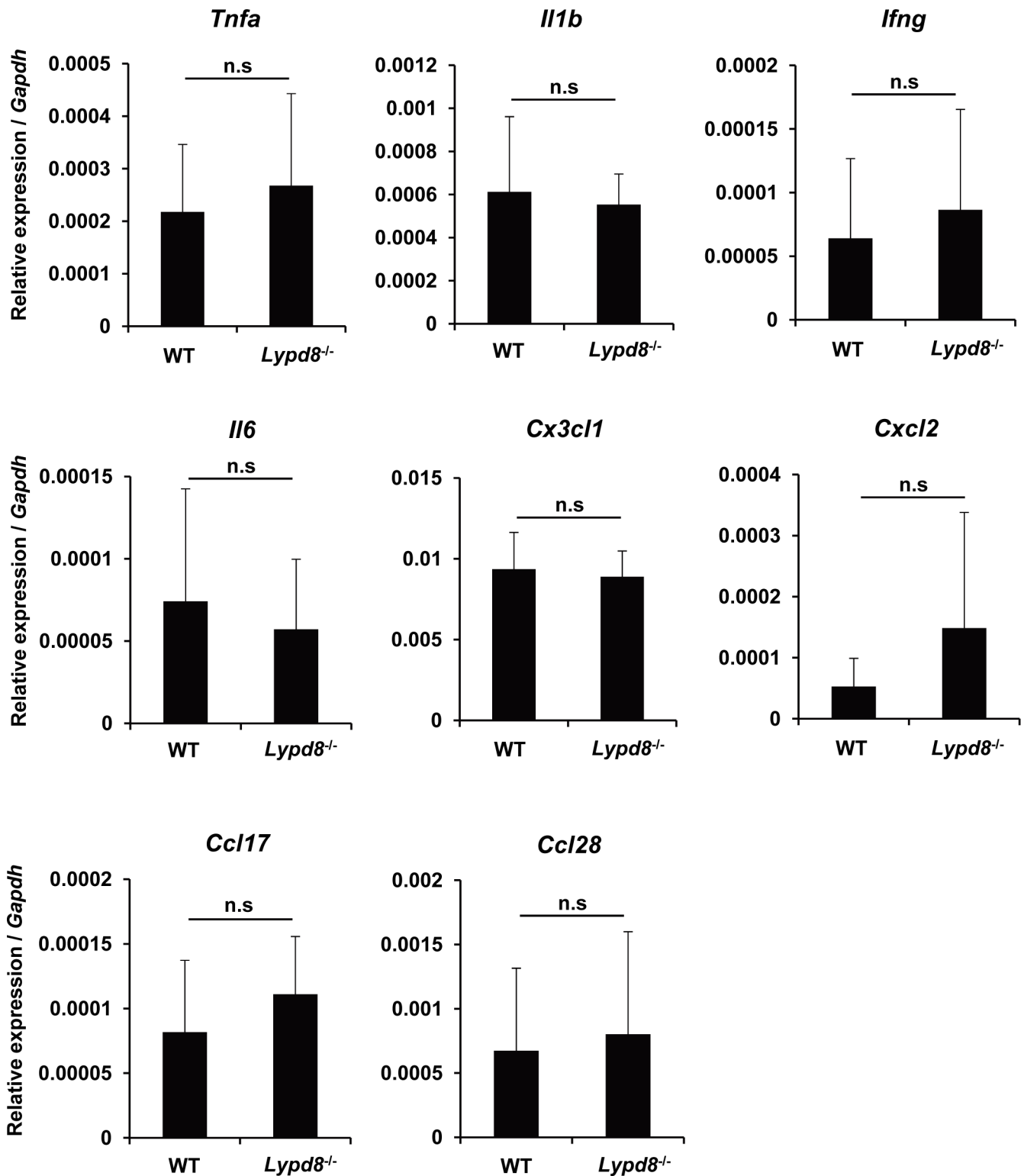
NcoI, separated by electrophoresis and hybridized with the radiolabelled probe shown in **a**. **c**, Quantitative RT-PCR analysis of *Lypd8* expression in the colon of wild-type and *Lypd8*^{-/-} mice. The values were normalized to *Gapdh* expression. Data are mean \pm s.d. ($n = 3$ per group).



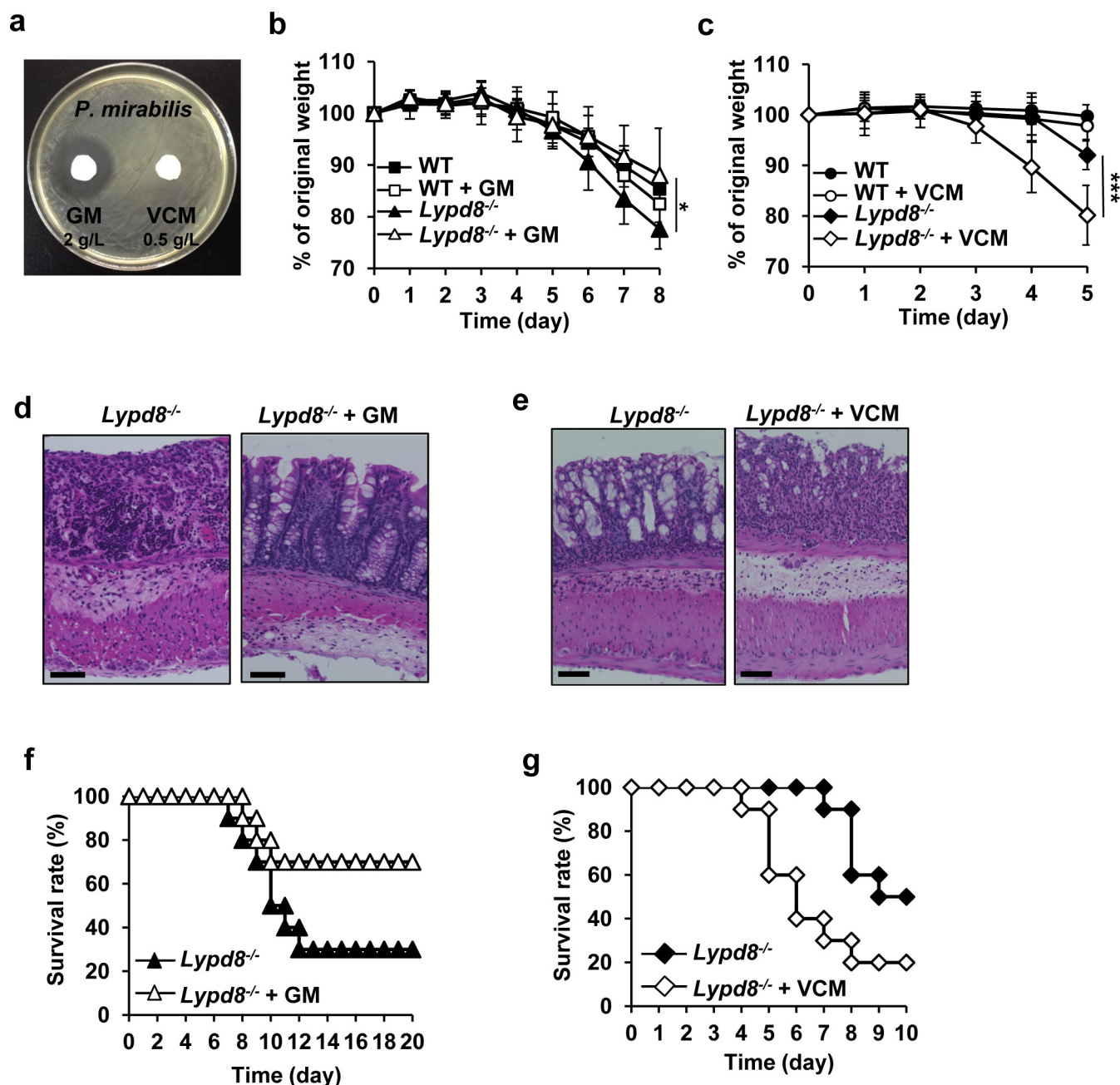
Extended Data Figure 4 | Bacteria are present in *Lypd8*^{-/-} epithelial cell layers. **a**, Principal component analysis of faecal microbiota in wild-type and *Lypd8*^{-/-} mice at genus level ($n = 8$ mice per genotype). **b**, Bacterial DNA isolated from luminal contents and PBS-washed colon



tissues of wild-type and *Lypd8*^{-/-} mice was analysed by quantitative PCR using universal primers for bacterial 16S rRNA genes. Data show the total bacterial DNA amounts compared to wild-type mice group, and are presented as mean \pm s.d. ($n = 6$ per group). ** $P < 0.01$.

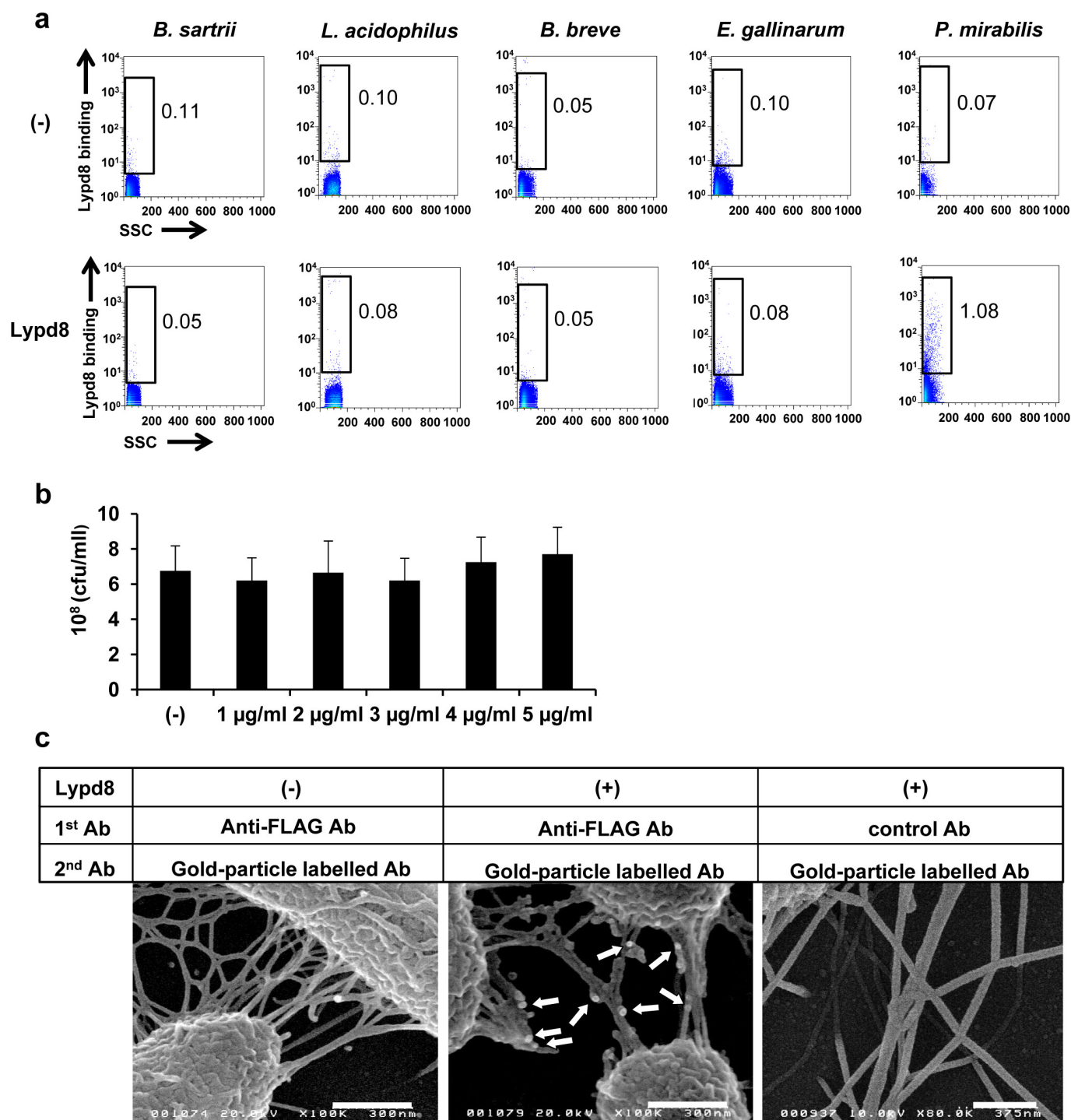


Extended Data Figure 5 | Inflammatory responses were not evident in *Lypd8*^{-/-} mice reared in an SPF facility. Expression of inflammatory cytokines and chemokines in the colon of wild-type and *Lypd8*^{-/-} mice. Data represent the mean \pm s.d. ($n = 10$ per group).



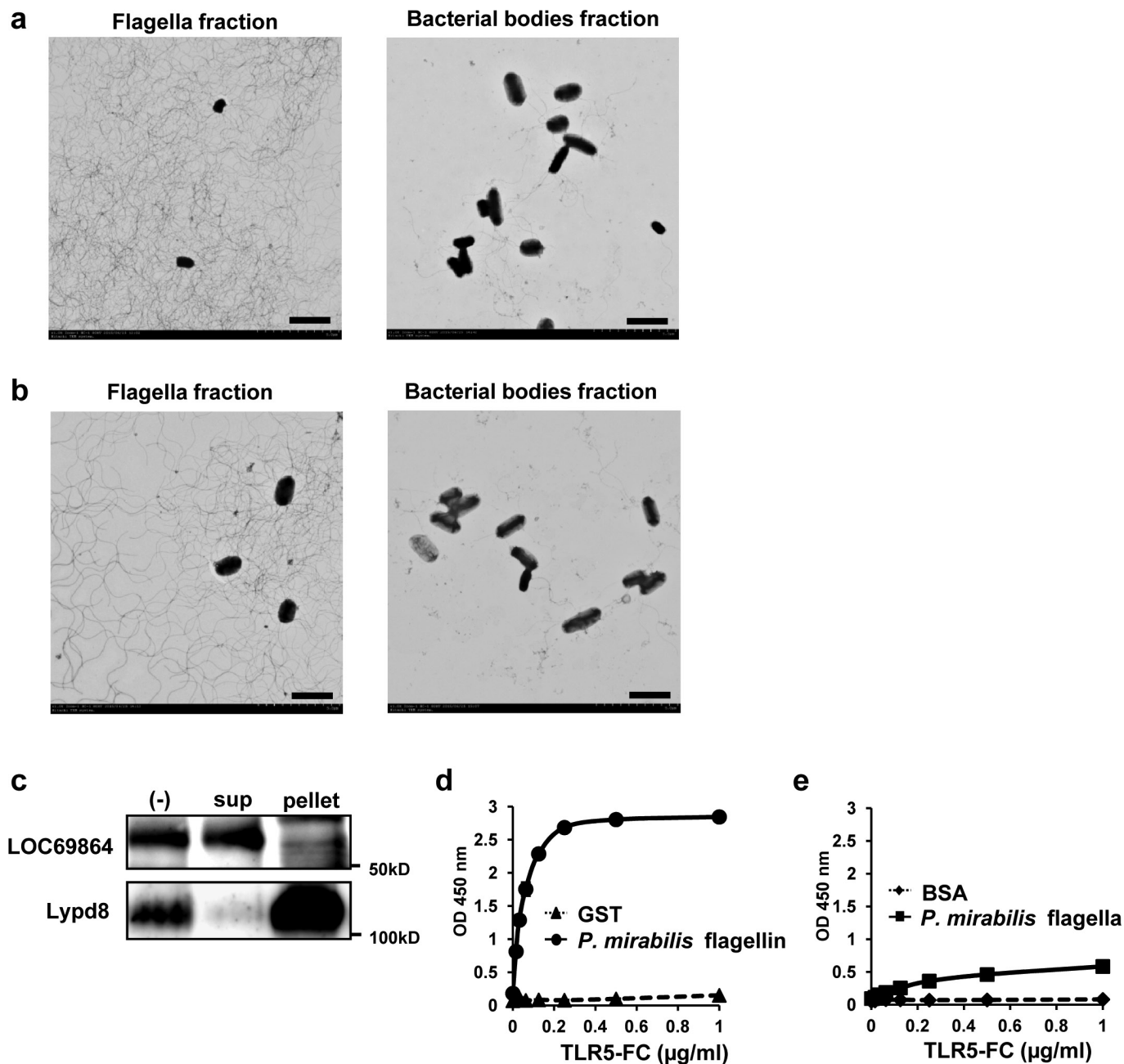
Extended Data Figure 6 | Gentamicin treatment ameliorated DSS-induced colitis in *Lypd8* mice. **a**, A paper disc method was performed to determine the sensitivity of gentamicin and vancomycin to *P. mirabilis*. **b–e**, Wild-type ($n=8$) and *Lypd8*^{-/-} mice ($n=8$) were orally administered gentamicin (**b, d**) or vancomycin (**c, e**) for 2 weeks and then administered DSS for 5 days. Percentage of weight change (**b, c**). H&E staining of colon tissues at day 8 (gentamicin) and day 5 (vancomycin) of DSS

administration (**d, e**). Scale bars, 50 μ m. Data are shown as mean \pm s.d. * $P < 0.05$, *** $P < 0.005$. **f**, *Lypd8*^{-/-} mice ($n=10$) and gentamicin-treated *Lypd8*^{-/-} mice ($n=10$) were orally administered 2% DSS for 5 days and their survival rate was monitored. **g**, *Lypd8*^{-/-} mice ($n=10$) and vancomycin-treated *Lypd8*^{-/-} mice ($n=10$) were orally administered 2% DSS for 5 days and their survival rate was monitored. Data show survival rate of each group.



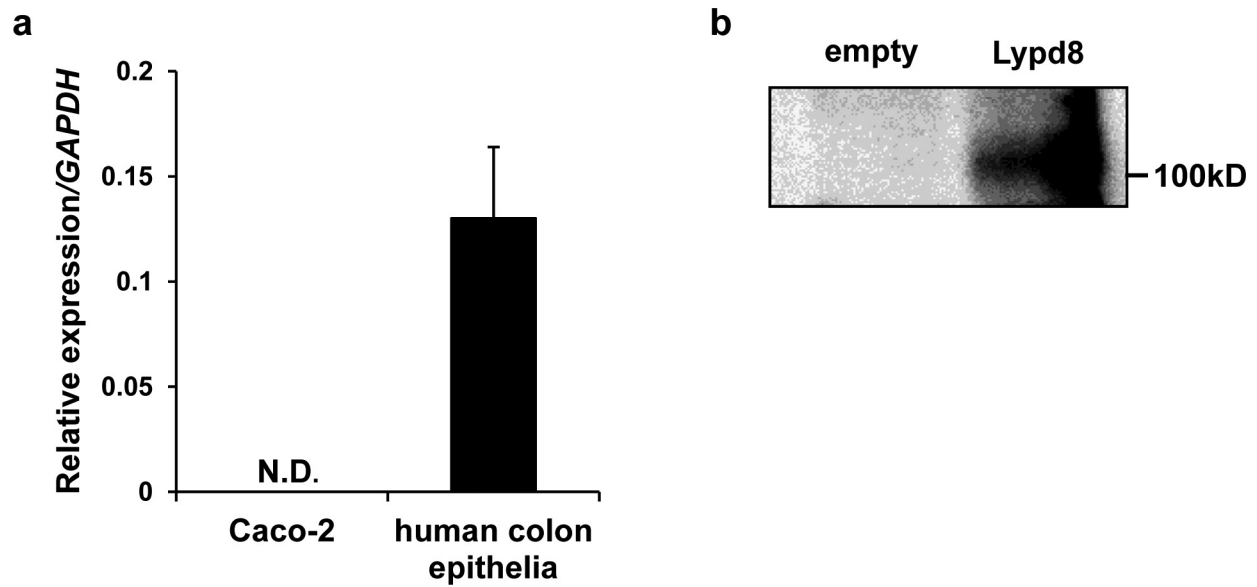
Extended Data Figure 7 | Lypd8 preferentially bound to *in-vitro*-cultured *P. mirabilis*. **a**, Representative dot plots of pure preparations of *in-vitro*-cultured *B. sartrii*, *L. acidophilus*, *B. breve*, *E. gallinarum* and *P. mirabilis* incubated with Flag-tagged Lypd8. Numbers within plots indicate percentages of bacteria bound by anti-Flag antibody in the respective areas. **b**, *P. mirabilis* was cultured in LB medium with the indicated concentrations of recombinant (r)Lypd8 protein for 6 h at

37°C. The mean CFU of cultured *P. mirabilis* are shown ($n = 4$ per group). Data are mean \pm s.d. **c**, *In-vitro*-cultured *P. mirabilis* were incubated with or without Flag-tagged Lypd8 and reacted with indicated 1st and 2nd antibodies. Scanning electron microscopic images of *P. mirabilis* labelled for immunogold detection of the bound Lypd8 were shown. White arrows indicate gold-particles. Scale bars, 300 nm.

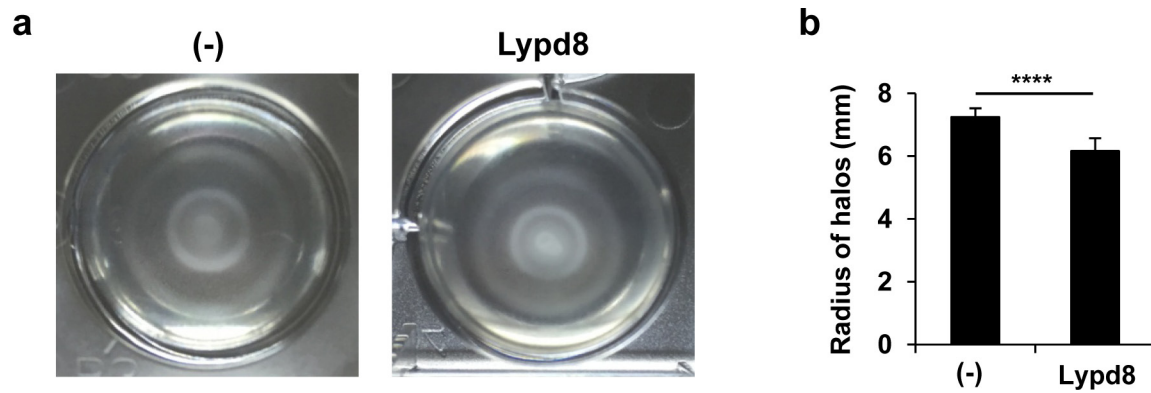


Extended Data Figure 8 | Isolation of flagella of *P. mirabilis* and *E. coli*. **a, b,** Bacterial flagella were isolated from bacterial suspension of *P. mirabilis* and *E. coli* by vigorous shaking and ultracentrifugation. Transmission electron microscopic images of flagella fraction and bacterial bodies fraction from *P. mirabilis* (**a**) and *E. coli* (**b**) are shown. Flagella and bacterial bodies were visualized by negative staining. Scale

bars, 2.5 μm . **c,** After incubation of Flag-tagged LOC69864 protein ($10 \text{ ng } \mu\text{l}^{-1}$) with flagella of *P. mirabilis* ($1 \text{ } \mu\text{g } \mu\text{l}^{-1}$), flagella were pelleted by ultracentrifugation. The supernatant (sup) and pellet suspension were separated by SDS-PAGE and immunoblotted with anti-Flag antibody. **d, e,** Association of mouse TLR5-Fc chimaera protein and flagellin (**d**) or flagella (**e**) of *P. mirabilis* was analysed by ELISA assay.



Extended Data Figure 9 | *LYPD8* expression in Caco-2 cells and human colon epithelia. **a**, Quantitative RT-PCR analysis of *LYPD8* expression in Caco-2 cells and human colonic epithelia isolated from normal parts of colonic tissues from colon cancer patients ($n = 3$). **b**, Immunoblot analysis for detecting Flag-tagged proteins in the supernatants on Caco-2 cells with or without Flag-tagged Lypd8.



Extended Data Figure 10 | Motility of *E. coli* was inhibited in semisolid agar containing Lypd8 protein. **a**, Representative images of *E. coli* motility in semisolid agar with or without Lypd8 protein 4 h after incubation. **b**, The radii of the motility halos by *E. coli* were measured at 4 h after incubation. Data are mean \pm s.d. ($n=6$ per group). **** $P < 0.001$.

Cerebral cavernous malformations arise from endothelial gain of MEKK3–KLF2/4 signalling

Zinan Zhou^{1*}, Alan T. Tang^{1*}, Weng-Yew Wong², Sharika Bamezai¹, Lauren M. Goddard¹, Robert Shenkar³, Su Zhou¹, Jisheng Yang¹, Alexander C. Wright⁴, Matthew Foley⁵, J. Simon C. Arthur⁶, Kevin J. Whitehead⁷, Issam A. Awad³, Dean Y. Li^{7,8}, Xiangjian Zheng^{2,9} & Mark L. Kahn¹

Cerebral cavernous malformations (CCMs) are common inherited and sporadic vascular malformations that cause strokes and seizures in younger individuals¹. CCMs arise from endothelial cell loss of KRIT1, CCM2 or PDCD10, non-homologous proteins that form an adaptor complex². How disruption of the CCM complex results in disease remains controversial, with numerous signalling pathways (including Rho^{3,4}, SMAD⁵ and Wnt/ β -catenin⁶) and processes such as endothelial–mesenchymal transition (EndMT)⁵ proposed to have causal roles. CCM2 binds to MEKK3 (refs 7–11), and we have recently shown that CCM complex regulation of MEKK3 is essential during vertebrate heart development¹². Here we investigate this mechanism in CCM disease pathogenesis. Using a neonatal mouse model of CCM disease, we show that expression of the MEKK3 target genes *Klf2* and *Klf4*, as well as Rho and ADAMTS protease activity, are increased in the endothelial cells of early CCM lesions. By contrast, we find no evidence of EndMT or increased SMAD or Wnt signalling during early CCM formation. Endothelial-specific

loss of *Map3k3* (also known as *Mekk3*), *Klf2* or *Klf4* markedly prevents lesion formation, reverses the increase in Rho activity, and rescues lethality. Consistent with these findings in mice, we show that endothelial expression of KLF2 and KLF4 is increased in human familial and sporadic CCM lesions, and that a disease-causing human CCM2 mutation abrogates the MEKK3 interaction without affecting CCM complex formation. These studies identify gain of MEKK3 signalling and KLF2/4 function as causal mechanisms for CCM pathogenesis that may be targeted to develop new CCM therapeutics.

To understand the cellular and molecular mechanisms that underlie CCM formation, we first examined the temporal course of lesion formation in mice with induced, endothelial-specific deletion of *Krit1* immediately after birth (iECre;*Krit1*^{fl/fl}, hereafter termed *Krit1*^{ECKO} mice). Vascular malformations were first detected at postnatal day 6 (P6) as dilated venules in the cerebellar white matter of *Krit1*^{ECKO} mice, with numerous mature lesions present by P11 (ref. 13) (Fig 1a–c and Extended Data Fig. 1).

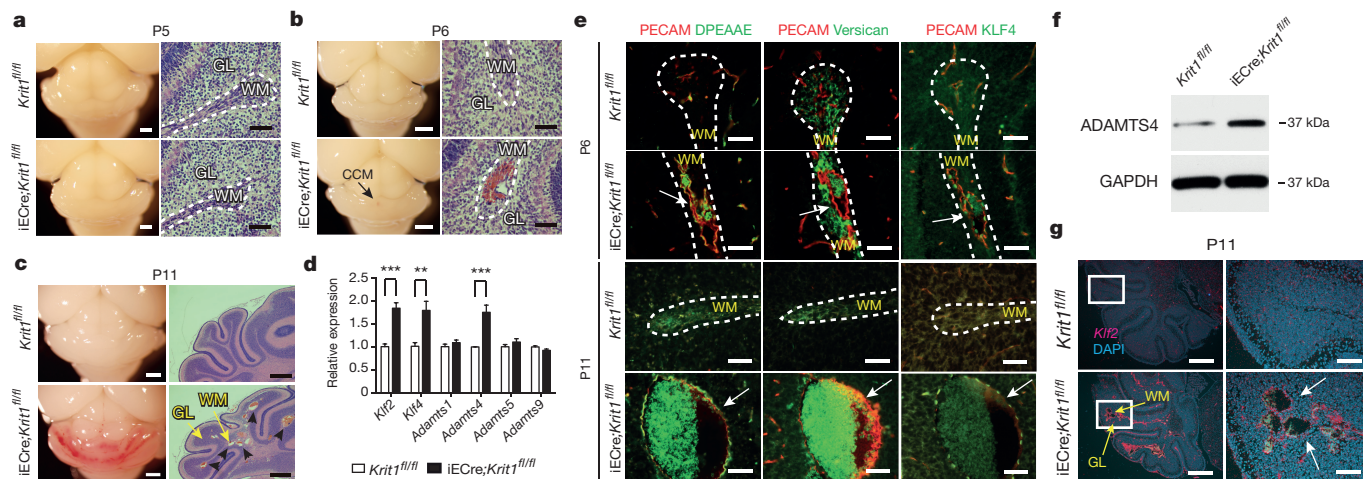


Figure 1 | KLF2, KLF4 and ADAMTS4 are increased in the earliest CCM lesions. **a, b**, Early CCM lesions (arrow) appear as haemorrhagic, dilated venules in the white matter (WM) of the cerebellum. GL, granular layer. Scale bars, 1 mm (white) and 50 μ m (black). **c**, Mature CCM lesions are detected as blood-filled caverns in the hindbrain of P11 *Krit1*^{ECKO} (iECre;*Krit1*^{fl/fl}) animals. Scale bars, 1 mm (white) and 500 μ m (black). **d**, Quantitative PCR (qPCR) analysis reveals increased *Klf2*, *Klf4* and *Adamts4* expression in hindbrain endothelial cells from *Krit1*^{ECKO} mice. ** $P < 0.01$; *** $P < 0.001$ (unpaired two-tailed Student's *t*-test). Error bars

indicate s.e.m. ($n = 4$). **e**, Abluminal ADAMTS-cleaved versican (DPEAAE) and nuclear KLF4 are detected in nascent (P6) and mature (P11) CCM lesions (arrows). Scale bars, 50 μ m. **f**, Increased levels of ADAMTS4 detected by immunoblotting of P11 *Krit1*^{ECKO} hindbrain lysate. Results are representative of 3 separate experiments. GAPDH immunoblot shown for loading control. **g**, Increased *Klf2* expression in CCM lesions (arrows) and meningeal vessels of *Krit1*^{ECKO} animals detected using *in situ* hybridization. Areas in boxes (left) are shown at higher magnification (right). Scale bars, 500 μ m (left) and 100 μ m (right). Dotted lines (**a, b, e**) outline cerebellar white matter.

¹Department of Medicine and Cardiovascular Institute, University of Pennsylvania, 3400 Civic Center Blvd, Philadelphia, Pennsylvania 19104, USA. ²Laboratory of Cardiovascular Signaling, Centenary Institute, Sydney, New South Wales 2050, Australia. ³Neurovascular Surgery Program, Section of Neurosurgery, Department of Surgery, The University of Chicago Medicine and Biological Sciences, Chicago, Illinois 60637, USA. ⁴Department of Radiology, University of Pennsylvania Medical Center, 3400 Spruce Street, Philadelphia, Pennsylvania 19104, USA. ⁵Sydney Microscopy & Microanalysis, University of Sydney, Sydney, New South Wales 2050, Australia. ⁶Division of Cell Signaling and Immunology, University of Dundee, Dundee DD1 5EH, UK. ⁷Division of Cardiovascular Medicine and the Program in Molecular Medicine, University of Utah, Salt Lake City, Utah 84112, USA. ⁸The Key Laboratory for Human Disease Gene Study of Sichuan Province, Institute of Laboratory Medicine, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, Sichuan 610072, China. ⁹Faculty of Medicine, Sydney Medical School, University of Sydney, Sydney, New South Wales 2050, Australia.

*These authors contributed equally to this work.

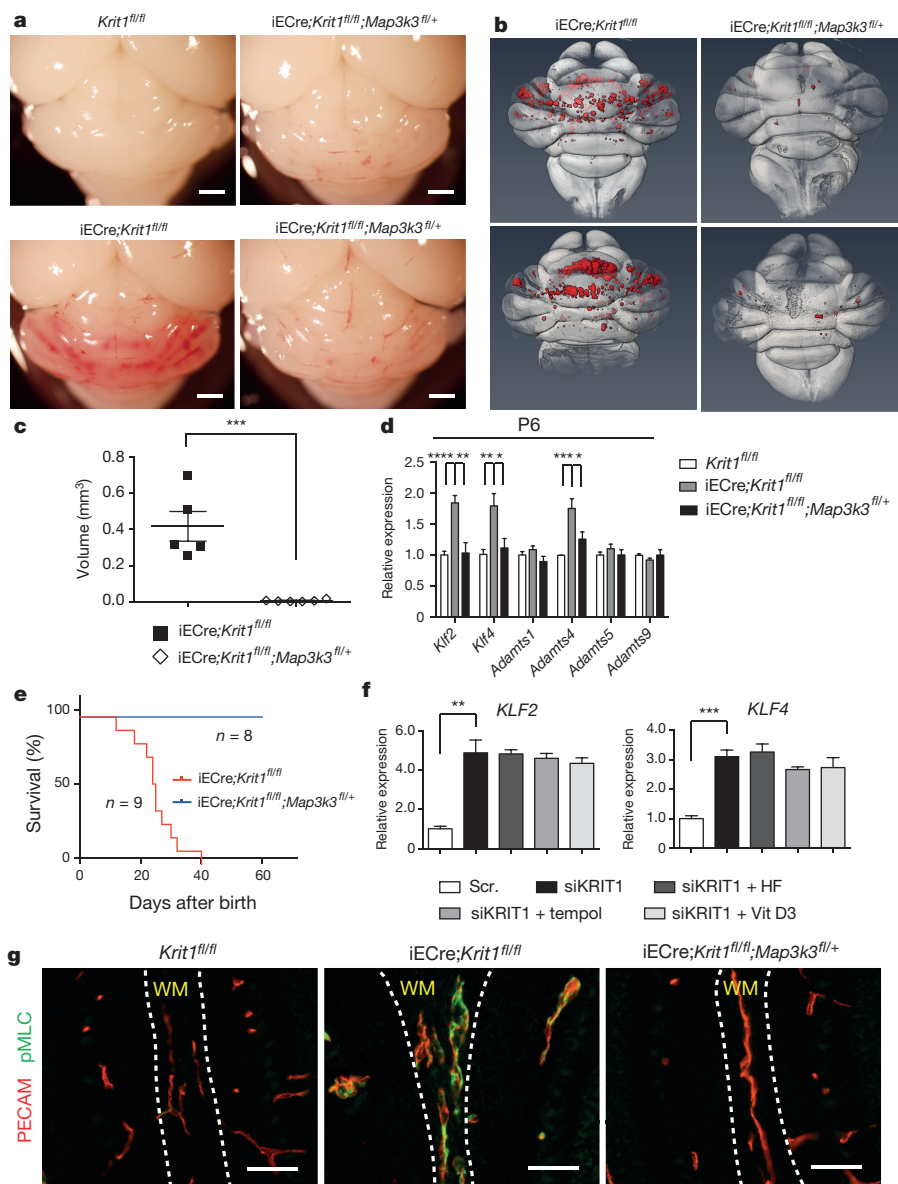


Figure 2 | Genetic rescue of CCM formation and increased Rho activity with loss of MEKK3. **a**, Visual detection of CCM lesions in the hindbrains of P11 *Krit1^{fl/fl}* (top left), *Krit1^{ECKO}* (*iECre;Krit1^{fl/fl}*; bottom left), and *Map3k3^{HetRSQ}* (*iECre;Krit1^{fl/fl};Map3k3^{fl/+}*; right) mice (right). Scale bars, 1 mm. **b**, Composite microCT images of *Krit1^{ECKO}* and *Map3k3^{HetRSQ}* hindbrains. CCM lesions are shown in red. **c**, Quantification of CCM lesion volumes ($n = 5$ for each group). **d**, qPCR analysis of gene expression in cerebellar endothelial cells ($n = 4$). **e**, Postnatal survival of *Krit1^{ECKO}* animals with and without endothelial loss of one *Map3k3* allele. $P = 0.0009$.

f, qPCR analysis of gene expression in human umbilical vein endothelial cells (HUVECs) treated with scrambled (Scr.) or *KRIT1*-targeting (siKRIT1) short interfering RNAs (siRNAs), alone and in the presence of the indicated Rho antagonists ($n = 3$). HF, hydroxyfasudil. **g**, Normalization of Rho activity with loss of MEKK3. PECAM and pMLC staining of white matter vessels in the indicated P6 littermate brains. Images in all panels are representative of 5 independent studies for each genotype. Scale bars, 50 μ m. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$ (unpaired two-tailed Student's *t*-test). Error bars indicate s.e.m.

We recently demonstrated that loss of KRIT1, CCM2 or PDCD10 in endothelial cells of the developing mouse heart upregulates expression of KLF2, KLF4, ADAMTS4 and ADAMTS5 owing to increased activity of the MEKK3 signalling pathway¹². Analysis of isolated cerebellar endothelial cells from neonatal *Krit1^{ECKO}* and littermate control mice at P6 revealed increased *Adamts4*, but not *Adamts5*, messenger RNA and protein, in addition to increased levels of both *Klf2* and *Klf4* (Fig. 1d, f). ADAMTS4 cleaves the proteoglycan versican to expose a neo-epitope (DPEAAE) that was detected immediately adjacent to the endothelial cells of both early and late CCM lesions (Fig. 1e). Increased levels of nuclear KLF4 protein and *Klf2* mRNA were also detected in the endothelial cells of CCM lesions and other vessels in the cerebellum (Fig. 1e, g). These findings reveal increased levels of KLF2, KLF4 and ADAMTS4 during the earliest phase of CCM lesion formation *in vivo*.

Recent studies have implicated numerous signalling mechanisms as causal for CCM formation, including the Rho^{3,4,14}, TGF- β /BMP⁵, Wnt/ β -catenin⁶ and Notch¹⁵ pathways. To determine whether changes in Rho activity are an early event in CCM pathogenesis, we examined the ROCK substrate phosphorylated-myosin light chain (pMLC). pMLC levels were markedly increased in the brain capillary and venous endothelial cells of P6 neonatal *Krit1^{ECKO}* mice compared with littermate controls, including those lining the earliest detectable CCM lesions (Extended Data Fig. 2a). TGF- β /BMP signalling through SMAD1 and SMAD3 and a mechanism of EndMT has been proposed to explain gene expression changes in KRIT1-deficient brain endothelial cells, including increased KLF4 (refs 5, 16). A recent study by the same group suggested that increased β -catenin signalling is a primary event that culminates in later EndMT⁶. To assess endothelial β -catenin

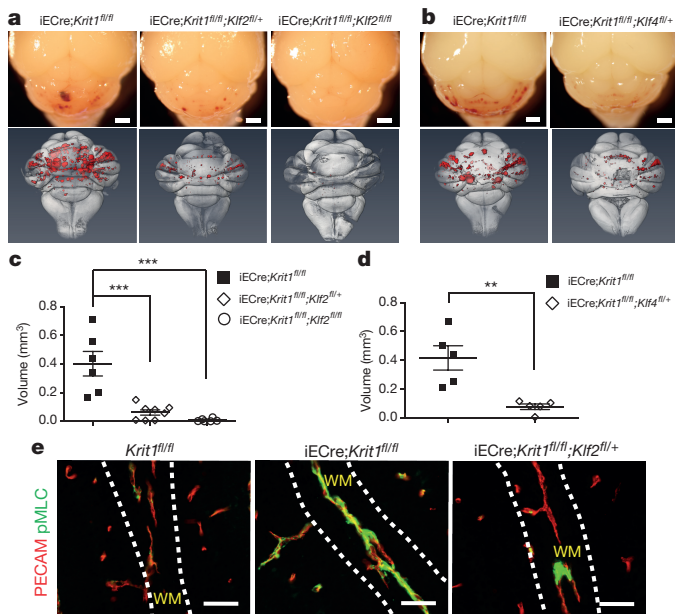


Figure 3 | Genetic rescue of CCM formation and increased Rho activity with loss of KLF2 and KLF4. **a, b**, Visual appearance of CCM lesions in *Krit1*^{ECKO} (iECre;*Krit1*^{fl/fl}), *Klf2*^{HetRSQ} (iECre;*Krit1*^{fl/fl};*Klf2*^{fl/fl}), *Klf2*^{HomoRSQ} (iECre;*Krit1*^{fl/fl};*Klf2*^{fl/fl}) and *Klf4*^{HetRSQ} (iECre;*Krit1*^{fl/fl};*Klf4*^{fl/fl}) mice (top), and composite microCT images of the same hindbrains (bottom). Scale bars, 1 mm. **c, d**, Volumetric quantification of CCM lesions in *Krit1*^{ECKO} littermates, with and without endothelial loss of one or two *Klf2* alleles (**c**) or one *Klf4* allele (**d**) ($n = 5-7$). $**P < 0.01$; $***P < 0.001$ (unpaired two-tailed Student's *t*-test). **e**, Rescue of increased Rho activity by loss of KLF2. PECAM and pMLC staining of white matter vessels in the indicated P6 littermate brains is shown. Images in all panels are representative of at least 4 independent studies for each genotype. Scale bars, 50 μ m.

signalling during CCM lesion formation, we generated neonatal *Krit1*^{ECKO} mice on a TCF/Lef:H2B-GFP Wnt/ β -catenin reporter background¹⁷. Immunostaining for green fluorescent protein (GFP) revealed β -catenin signalling in white matter vascular endothelium that was not increased in P6 CCM lesions (Extended Data Fig. 2b), while the levels of both GFP protein and the β -catenin target genes *Axin2* and *Lef1* were unchanged in cerebellar endothelial cells isolated from P6 or P11 neonatal *Krit1*^{ECKO} mice (Extended Data Fig. 2c, d). Immunostaining of brain sections and immunoblotting of cerebellar endothelial cell protein also revealed no changes in pSMAD3 (Extended Data Fig. 2e, f). Finally, analysis of cerebellar endothelial gene expression analysis revealed no change in the expression of Notch target genes at P6, although an increase in *Hes1* was noted at P11 (Extended Data Fig. 3). These studies reveal that primary CCM lesion formation is associated with increases in *Klf2*, *Klf4* and *Adams4* expression and Rho/ROCK activity, but not in TGF- β /BMP, Wnt/ β -catenin or Notch signalling.

The above findings suggested that changes in KLF2/4 and ADAMTS4 expression may be causal for CCM formation. The CCM complex directly binds MEKK3 (refs 7–11), a MAP3 kinase known to regulate KLF2 and KLF4 expression in the embryonic heart in cultured endothelial cells¹², and we previously found that *Map3k3* haploinsufficiency rescues the loss of cardiac jelly associated with endocardial loss of CCM signalling¹². *Map3k3* haploinsufficiency was also found to rescue the early embryonic lethality conferred by pan-endothelial loss of KRIT1 (Extended Data Fig. 4a and ref. 18), suggesting that excess endothelial MEKK3 signalling may have a broad role in the cardiovascular phenotypes associated with loss of CCM signalling.

To determine whether this model underlies CCM formation, we generated iECre;*Krit1*^{fl/fl};*Map3k3*^{HetRSQ} mice (termed *Map3k3*^{HetRSQ} mice). Visual inspection of the hindbrains of P11 *Map3k3*^{HetRSQ} mice compared with neonatal *Krit1*^{ECKO} littermate controls revealed a

notable reduction in the number and size of vascular lesions (Fig. 2a). To quantify CCM formation, we imaged P11 hindbrains using contrast-enhanced, high resolution X-ray micro-computed tomography (microCT), and measured actual lesion volumes using semi-automated software. *Map3k3*^{HetRSQ} mouse hindbrains exhibited nearly complete prevention of the lesion phenotype compared with neonatal *Krit1*^{ECKO} littermates, as assessed by hindbrain microCT imaging (Fig. 2b), blinded measurement of total CCM lesion volume (Fig. 2c), and measurements of *Klf2*, *Klf4* and *Adams4* in P6 cerebellar endothelial cells (Fig. 2d). While almost all neonatal *Krit1*^{ECKO} mice were dead by P30, all *Map3k3*^{HetRSQ} animals remained alive (Fig. 2e), and exhibited normal growth and development. Finally, partial loss of MEKK3 also fully rescued CCM lesion formation in *Ccm2*^{ECKO} (iECre;*Ccm2*^{fl/fl}) mice (Extended Data Fig. 4b). These genetic findings support the conclusion that CCM lesions arise from gain of MEKK3 signalling and altered downstream gene expression in the endothelium.

Increased endothelial pMLC is coincident with increased *Klf2*, *Klf4* and *Adams4* expression in the earliest CCM lesions (Fig. 1e, g and Extended Data Fig. 2a), suggesting either that changes in Rho/ROCK activity are downstream of changes in MEKK3 activity or vice versa. The Rho-inhibiting agents hydroxyfasudil, tempol and vitamin D3 (ref. 19) failed to reverse the increase in *KLF2* and *KLF4* expression conferred by loss of KRIT1 in cultured human endothelial cells (Fig. 2f), suggesting that Rho is not upstream of the KLF2/4 expression changes associated with loss of CCM function. By contrast, P6 *Map3k3*^{HetRSQ} mice exhibited a complete normalization of endothelial pMLC staining (Fig. 2g), indicating that increased Rho activity arises secondary to increased MEKK3 signalling during CCM formation.

To test the roles of KLF2 and KLF4 in CCM pathogenesis, we measured lesion formation in *Klf2*^{HetRSQ} (iECre;*Krit1*^{fl/fl};*Klf2*^{fl/fl}), *Klf2*^{HomoRSQ} (iECre;*Krit1*^{fl/fl};*Klf2*^{fl/fl}) and *Klf4*^{HetRSQ} (iECre;*Krit1*^{fl/fl};*Klf4*^{fl/fl}) mice compared with littermate neonatal *Krit1*^{ECKO} controls at P11. *Klf2*^{HetRSQ} and *Klf4*^{HetRSQ} mice exhibited a marked but incomplete prevention of lesion formation (80% and 75% rescue for *Klf2*^{HetRSQ} and *Klf4*^{HetRSQ} mice, respectively) based on visual inspection of cerebellar lesions and quantification of CCM lesion volume after microCT imaging (Fig. 3a–d). Remarkably, *Klf2*^{HomoRSQ} mice exhibited 99% rescue (Fig. 3a, c), with only a small amount of venule dilatation visible histologically (Extended Data Fig. 5). pMLC staining was normalized in P6 *Klf2*^{HetRSQ} mice (Fig. 3e), indicating that increased Rho/ROCK activity arises secondary to increased KLF2 expression. These findings identify gain of KLF2 and KLF4 function as causal for CCM formation, and suggest that these transcription factors are the primary downstream targets of MEKK3 in this disease model. They also highlight the notable molecular conservation of this endothelial pathway: from zebrafish to mammals, and from embryonic vascular endothelium and endocardium to postnatal brain endothelium.

To determine whether MEKK3–KLF2/4 signalling is increased in human CCMs, we examined resected lesions from two patients with familial CCM bearing *KRIT1* and *PDCD10* germline mutations, and two sporadic CCM patients lacking any previous genetic or molecular data. Markedly increased nuclear KLF2 and KLF4 expression was observed in the endothelial cells of both familial and sporadic human CCM lesions (Fig. 4a, b), consistent with increased MEKK3 signalling and studies performed using the mouse model. MEKK3 binds CCM2 through the carboxy-terminal helical harmonin domain of CCM2, and CCM2 truncation mutants lacking this domain do not bind MEKK3 (Extended Data Fig. 6a, b and refs 10, 11, 20, 21). A literature search identified a familial CCM patient with a four-nucleotide duplication in the last exon of CCM2 (CCCTdup) predicted to delete most of the helical harmonin domain²² (Fig. 4c). CCM2 CCCTdup expressed normally in HEK293T cells and bound KRIT1 and PDCD10 in a manner indistinguishable from wild-type CCM2, but failed to interact with MEKK3 (Fig. 4c–e and Extended Data Fig. 6c). These results suggest that specific disruption of the CCM2–MEKK3 interaction is sufficient to cause familial CCM

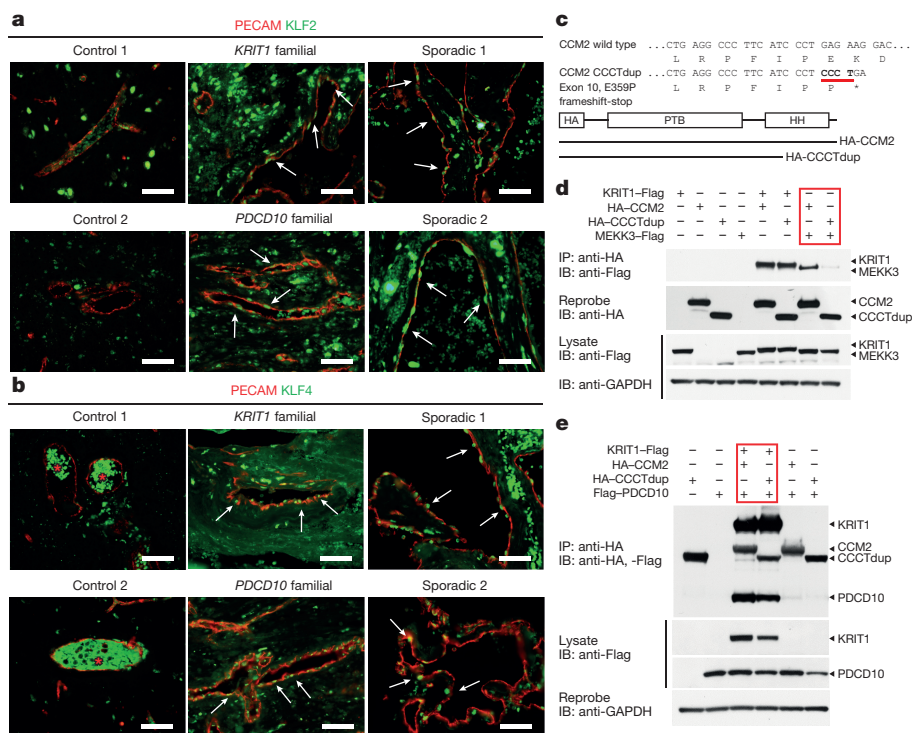


Figure 4 | Human CCMs exhibit high levels of endothelial KLF2 and KLF4 and arise owing to selective loss of CCM2-MEKK3 interaction. **a, b**, Immunostaining for PECAM plus either KLF2 (**a**) or KLF4 (**b**) is shown for cerebral vessels in individuals without CCM disease (control; left), for CCM lesions arising owing to germline mutations in *KRIT1* or *PDCD10* (familial; middle), and for sporadic CCM lesions from two individuals (sporadic; right). Arrows indicate nuclear KLF2-high and KLF4-high endothelial cells in the CCM lesions. Red asterisks indicate fluorescence due to trapped intravascular erythrocytes. Scale bars, 50 μ m.

c, Schematic representation of the CCM2 CCCT duplication (CCCTdup). HH, helical harmonin domain; PTB, phosphotyrosine-binding domain. **d**, CCM2 CCCTdup does not bind MEKK3. The indicated proteins were expressed in HEK293T cells, immunoprecipitated (IP) using anti-haemagglutinin (HA) antibodies, and immunoblotted (IB) using the antibodies shown. **e**, CCM2 CCCTdup binds KRIT1 and PDGCD10 like wild-type CCM2. The indicated proteins were expressed and co-immunoprecipitated as described in **c**. Results shown are representative of at least 3 separate experiments.

disease, and that human CCMs also arise due to loss of MEKK3 regulation and increased expression of KLF2 and KLF4.

How gain of MEKK3-KLF2/4 signalling confers CCM formation is unclear. It has been proposed that EndMT underlies CCM pathogenesis^{5,6,23}, but we detect no evidence of a change in phenotype from endothelial to mesenchymal with loss of CCM signalling, and demonstrate that the loss of the non-mesenchymal transcription factor KLF2 is sufficient to rescue CCM formation fully. Our studies identify two effector pathways downstream of MEKK3-KLF2/4 signalling that may drive early CCM pathogenesis: Rho signalling and ADAMTS proteolytic activity. Increased Rho activity has been linked to increased stress fibre formation, loosened junctions and decreased tube formation in cultured endothelial cells^{3,14,24}, and loss of vascular integrity in mice^{4,25}, but whether and how these changes might cause CCM formation is not yet clear. Increased ADAMTS activity is conferred by gain of KLF2 and KLF4 function (Extended Data Fig. 7), and may confer CCM formation through breakdown of a proteoglycan matrix that is required specifically for the CNS vasculature (for example, versican; Extended Data Fig. 1b). A proteolytic mechanism would also explain the autosomal dominant inheritance of this disease, as CCM-deficient endothelial cells generated by a rare second-hit mechanism^{26,27} could degrade the matrix supporting an entire vessel, thereby creating a cavernous malformation.

Finally, these studies may help to identify new therapies for CCM disease. Since MEKK3 upregulates KLF2 and KLF4 through the MEK5 and ERK5 downstream MAPKs^{12,28}, we evaluated the effects of available inhibitors on CCM formation. BIX02189 (anti-MEK5) and XMD17-109 (anti-ERK5) reversed the increase in KLF2 and KLF4 expression associated with loss of CCM signalling in cultured endothelial cells (Extended Data Fig. 8), but failed to affect CCM formation at the low doses tolerated by neonatal *Krit1*^{ECKO} mice (not shown).

Therapies targeting Rho (for example, fasudil²⁹ or other recently identified agents¹⁹) or the ADAMTS proteases may prove more effective, but future studies that rigorously test the causal role of these putative downstream effectors remain essential for the rational development of CCM disease therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 August 2015; accepted 27 January 2016.

Published online 30 March 2016.

- Cavalcanti, D. D. *et al.* Cerebral cavernous malformations: from genes to proteins to disease. *J. Neurosurg.* **116**, 122–132 (2012).
- Plummer, N. W., Zawistowski, J. S. & Marchuk, D. A. Genetics of cerebral cavernous malformations. *Curr. Neurol. Neurosci. Rep.* **5**, 391–396 (2005).
- Whitehead, K. J. *et al.* The cerebral cavernous malformation signaling pathway promotes vascular integrity via Rho GTPases. *Nature Med.* **15**, 177–184 (2009).
- Stockton, R. A., Shenkar, R., Awad, I. A. & Ginsberg, M. H. Cerebral cavernous malformations proteins inhibit Rho kinase to stabilize vascular integrity. *J. Exp. Med.* **207**, 881–896 (2010).
- Maddaluno, L. *et al.* EndMT contributes to the onset and progression of cerebral cavernous malformations. *Nature* **498**, 492–496 (2013).
- Bravi, L. *et al.* Sulindac metabolites decrease cerebrovascular malformations in CCM3-knockout mice. *Proc. Natl Acad. Sci. USA* **112**, 8421–8426 (2015).
- Uhlik, M. T. *et al.* Rac-MEKK3-MKK3 scaffolding for p38 MAPK activation during hyperosmotic shock. *Nature Cell Biol.* **5**, 1104–1110 (2003).
- Zawistowski, J. S. *et al.* CCM1 and CCM2 protein interactions in cell signaling: implications for cerebral cavernous malformations pathogenesis. *Hum. Mol. Genet.* **14**, 2521–2531 (2005).
- Cullere, X., Plovie, E., Bennett, P. M., MacRae, C. A. & Mayadas, T. N. The cerebral cavernous malformation proteins CCM2L and CCM2 prevent the activation of the MAP kinase MEKK3. *Proc. Natl Acad. Sci. USA* (2015).
- Fisher, O. S. *et al.* Structure and vascular function of MEKK3-cerebral cavernous malformations 2 complex. *Nature Commun.* **6**, 7937 (2015).

11. Wang, X. *et al.* Structural insights into the molecular recognition between cerebral cavernous malformation 2 and mitogen-activated protein kinase kinase 3. *Structure* **6**, 1087–1096 (2015).
12. Zhou, Z. *et al.* The cerebral cavernous malformation pathway controls cardiac development via regulation of endocardial MEKK3 signaling and KLF expression. *Dev. Cell* **32**, 168–180 (2015).
13. Boulday, G. *et al.* Developmental timing of CCM2 loss influences cerebral cavernous malformations in mice. *J. Exp. Med.* **208**, 1835–1847 (2011).
14. Glading, A., Han, J., Stockton, R. A. & Ginsberg, M. H. KRIT-1/CCM1 is a Rap1 effector that regulates endothelial cell cell junctions. *J. Cell Biol.* **179**, 247–254 (2007).
15. Wüsthube, J. *et al.* Cerebral cavernous malformation protein CCM1 inhibits sprouting angiogenesis by activating DELTA-NOTCH signaling. *Proc. Natl Acad. Sci. USA* **107**, 12640–12645 (2010).
16. Cuttano, R. *et al.* KLF4 is a key determinant in the development and progression of cerebral cavernous malformations. *EMBO Mol. Med.* **8**, 6–24 (2015).
17. Ferrer-Vaquero, A. *et al.* A sensitive and bright single-cell resolution live imaging reporter of Wnt/ β -catenin signaling in the mouse. *BMC Dev. Biol.* **10**, 121 (2010).
18. Whitehead, K. J., Plummer, N. W., Adams, J. A., Marchuk, D. A. & Li, D. Y. Ccm1 is required for arterial morphogenesis: implications for the etiology of human cavernous malformations. *Development* **131**, 1437–1448 (2004).
19. Gibson, C. C. *et al.* Strategy for identifying repurposed drugs for the treatment of cerebral cavernous malformation. *Circulation* **131**, 289–299 (2015).
20. Fisher, O. S. *et al.* Structural basis for the disruption of the cerebral cavernous malformations 2 (CCM2) interaction with Krev interaction trapped 1 (KRIT1) by disease-associated mutations. *J. Biol. Chem.* **290**, 2842–2853 (2015).
21. Draheim, K. M. *et al.* CCM2–CCM3 interaction stabilizes their protein expression and permits endothelial network formation. *J. Cell Biol.* **208**, 987–1001 (2015).
22. Spiegler, S. *et al.* High mutation detection rates in cerebral cavernous malformation upon stringent inclusion criteria: one-third of probands are minors. *Mol. Genet. Genomic Med.* **2**, 176–185 (2014).
23. Marchi, S. *et al.* Defective autophagy is a key feature of cerebral cavernous malformations. *EMBO Mol. Med.* **7**, 1403–1417 (2015).
24. Borikova, A. L. *et al.* Rho kinase inhibition rescues the endothelial cell cerebral cavernous malformation phenotype. *J. Biol. Chem.* **285**, 11760–11764 (2010).
25. Zheng, X. *et al.* Dynamic regulation of the cerebral cavernous malformation pathway controls vascular stability and growth. *Dev. Cell* **23**, 342–355 (2012).
26. Pagenstecher, A., Stahl, S., Sure, U. & Felbor, U. A two-hit mechanism causes cerebral cavernous malformations: complete inactivation of CCM1, CCM2 or CCM3 in affected endothelial cells. *Hum. Mol. Genet.* **18**, 911–918 (2009).
27. Akers, A. L., Johnson, E., Steinberg, G. K., Zabramski, J. M. & Marchuk, D. A. Biallelic somatic and germline mutations in cerebral cavernous malformations (CCMs): evidence for a two-hit mechanism of CCM pathogenesis. *Hum. Mol. Genet.* **18**, 919–930 (2009).
28. Chao, T. H., Hayashi, M., Tapping, R. I., Kato, Y. & Lee, J. D. MEKK3 directly regulates MEK5 activity as part of the big mitogen-activated protein kinase 1 (BMK1) signaling pathway. *J. Biol. Chem.* **274**, 36035–36038 (1999).
29. McDonald, D. A. *et al.* Fasudil decreases lesion burden in a murine model of cerebral cavernous malformation disease. *Stroke* **43**, 571–574 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the Kahn laboratory for their comments during the course of this work. These studies were supported by National Institutes of Health (NIH) grants R01HL094326 (to M.L.K.), P01NS092521 (to I.A.A. and M.L.K.), VAMC 1101BX002976 (to D.Y.L.), R01HL-084516 (to D.Y.L.), R01NS075168 (to K.J.W.), T32HL07439 (to A.T.T.), and Australian NHMRC project grant 161558 (to X.Z.).

Author Contributions Z.Z., A.T.T. and W.-Y.W. designed and performed most of the experiments and helped to write the manuscript. S.B. performed all of the microCT CCM lesion measurements in a blinded manner. J.S.C.A., K.J.W., R.S., I.A.A. and D.Y.L. provided critical reagents. S.B., L.M.G., S.Z., J.Y., A.C.W., M.F., X.Z. and M.L.K. helped to design and perform the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.L.K. (markkahn@mail.med.upenn.edu).

METHODS

Mice. VE-cadherin CreERT2 ('iECre') transgenic mice were a gift from R. Adams³⁰. *Tie2^{Cre}*, *Krt1^{fl/fl}*, *Ccm2^{fl/fl}*, *Map3k3^{fl/fl}*, *Klf2^{fl/fl}* and *Klf4^{fl/fl}* animals have been previously described^{12,25,31–34}. The TCF/LEF:H2B-GFP reporter line was obtained from Jackson Laboratories (013752)¹⁷. All intercrossed animals were maintained on mixed-strain backgrounds. Breeding pairs between 2 and 8 months of age were used to generate the neonatal CCM mouse model pups. Male and female animals were used in equal numbers for all experiments. The University of Pennsylvania Institutional Animal Care and Use Committee approved all animal protocols, and all procedures were performed in accordance with these protocols.

Induction of the neonatal CCM mouse model. Pups were intragastrically injected with 25 µg 4-hydroxytamoxifen (Sigma Aldrich H7904) at intervals of 1, 3 and 5 days after birth. The 4-hydroxytamoxifen was dissolved in warm 10% ethanol/corn oil vehicle, and 50 µl total volume was used per injection and delivered with a 30-gauge needle. Importantly, pups were injected in a blinded fashion without knowledge of genotypes, and there were no phenotypic indications within the first 5 days after birth. Pups were collected at specified time points 6 or 11 days after birth.

Mouse embryos for branchial arch artery rescue. Embryos were collected at embryonic day 10.5 (E10.5), gross images were taken, and embryos were placed in 4% paraformaldehyde overnight. Samples were processed for histology as described.

Histology. Embryos or mouse brains were fixed in 4% formaldehyde overnight, dehydrated in 100% ethanol, and embedded in paraffin. Sections (8-µm thick) were used for immunohistochemistry and haematoxylin and eosin staining. *Klf2* *in situ* hybridization was performed as previously reported³². The following antibodies were used for immunostaining: rat anti-PECAM (1:500, R&D MAB3628), rat anti-PECAM (1:20, Histo Bio Tech DIA-310), rabbit anti-versican (1:200, Millipore AB1033), rabbit anti-DPEAAE (1:200, Pierce-Antibodies PA1-1748A), mouse anti-KLF4 (1:100, R&D AF3158), rabbit anti-pSMAD3 (1:25, Abcam ab52903), rabbit anti-pMLC2 (1:200, Cell Signaling 3674S).

Isolation of cerebellar endothelial cells. Cerebellar endothelial cells were isolated by enzymatic digestion followed by separation using magnetic-activated cell sorting (MACS MS system, Miltenyi Biotec). Mice were first anaesthetized with Avertin (Sigma Aldrich T48402) and perfused with sterile PBS. Cerebella of the mice were then digested by 1 mg ml⁻¹ collagenase/dispase (Sigma) and 0.02 mg ml⁻¹ DNase I (Sigma) in complete DMEM for 10 min at 37°C with gentle shaking. The digestion was then passed through a 70-µm cell strainer (BD Biosciences). Cells were centrifuged, resuspended and incubated with anti-mouse CD31-antibody-conjugated microbeads for 15 min at 4°C. Microbead-bound cells were then washed and separated using MACS MS columns according to vendor protocol. Cells bound to the magnetic column were eluted and centrifuged for downstream applications, including western blotting and qPCR analysis.

Western blotting of protein from isolated endothelial cells. Protein of freshly isolated cerebellar endothelial cells was purified using RIPA buffer with complete protease inhibitor cocktail (Roche) and PhosSTOP phosphatase inhibitor cocktail (Roche). The following antibodies were used for immunoblotting: rabbit anti-GAPDH (1:5,000, Cell Signaling 2118), rabbit anti-ADAMTS4 (1:1,000, Abcam ab28285), rabbit anti-SMAD3 (1:1,000, Cell Signaling 9513), rabbit anti-pSMAD3 (1:1,000, Cell Signaling C25A9), rabbit anti-GFP (1:1,000, Abcam ab290).

Gene expression analysis. Total RNA of isolated endothelial cells was isolated with that RNeasy micro kit (Qiagen). For qPCR analysis, cDNA was synthesized from 150 ng total RNA using the Superscript III Reverse Transcriptase (Invitrogen). Real-time PCR was performed with Power SYBR Green PCR Master Mix (Applied Biosciences) using the following mouse primers: *Gapdh* forward 5'-AAATGGTGAAGTCCGGTGTGAACG-3', *Gapdh* reverse 5'-ATCTCCACTTTGCCACTGC-3'; *Klf2* forward 5'-CGCCTCGGGTTCATTTC-3', *Klf2* reverse 5'-AGCCTATCTTCCGTCCTTT-3'; *Klf4* forward 5'-GTGCCCGGACTAACCGTTG-3', *Klf4* reverse 5'-GTCGTTGAACCTCTCGGTCT-3'; *Adams1* forward 5'-CTCTACCCCTTCGGAATTTCTG-3', *Adams1* reverse 5'-GGAGCCACATAAATCCTGTCTG-3'; *Adams4* forward 5'-CAGTGCCCGATTCATCACT-3', *Adams4* reverse 5'-GAGTCAGGACCGAAGGTCAG-3'; *Adams5* forward 5'-CGACCCCTCAAGAACTTTTGC-3', *Adams5* reverse 5'-CGTCATGAGAAAGGCCAAGT-3'; *Adams9* forward 5'-TTGGGACCTGCTCAAGAAG-3', *Adams9* reverse 5'-ACCATTGATGTTGAAGTGTTC-3'; *Axin2* forward 5'-CAGCCCTGTGGTTCAAGCT-3', *Axin2* reverse 5'-GGTAGATTCTGATGGCCGATGT-3'; *Lef1* forward 5'-TAACGAGTCCGAATCATCCAGC-3', *Lef1* reverse 5'-TTCATCAGG GTGTTCTCTGGCCTT-3'; *Hes1* forward 5'-AAGCCTATCATGGAGAAGAGGCG-3', *Hes1* reverse 5'-GGAATGCCGGGAGCTATCTTTCTT-3'; *Hey1* forward 5'-GCGCGGACGAGAATGGAAA-3', *Hey1* reverse 5'-TCAGGTGATCCACAGTCATGTG-3'; *Hey2* forward 5'-AAGCGCCCTGTGAGGAAAC-3', *Hey2* reverse 5'-GGTAGTTGTCGGTGAATTGGAC-3'.

X-ray microCT-based quantification of neonatal, cerebellar lesions. Eleven-day-old pups were anaesthetized with Avertin and underwent intracardiac perfusion with PBS and 2% paraformaldehyde. Brains were quickly dissected and fixed overnight in a 2.5% glutaraldehyde, 4% formaldehyde, 0.1 M sodium phosphate buffer. Then, brains were rinsed several times with 0.1 M sodium phosphate buffer and hindbrains were detached from mid/forebrain structures by severing the pons. Next, hindbrains were soaked overnight in 2% osmium tetroxide solution (tissue contrast agent) and washed several times with water.

Hindbrains were randomized and scanned by a blinded operator using Xradia MicroCT system (Xradia MicroXCT-400). Images were acquired at 50 kV, 10 W, 721 projections, 3-s integration per 180° rotation.

Raw image stacks from each scanned hindbrain were analysed using Avizo 3D image processing software (Lite edition, FEI Visualization Sciences Group). Each image stack was imported and lesions were labelled in semi-automated, blinded fashion using a region-growing segmentation algorithm included with the software. Labelled lesions could then be analysed for pixel volume and converted to cubic micrometres. Importantly, to prevent measurement bias, a single individual, without direct experimental involvement or knowledge of genotypes, was used to label lesions of randomly ordered brains. This individual was trained before beginning analysis of the hindbrains. New-user accuracy was confirmed by comparison of results from a test brain analysis with the results from an experienced user. A flow-chart of this process is shown in Extended Data Fig. 9.

Post-lesion labelling, the hindbrain image stack was volume rendered and overlaid with the labelled lesions in the Avizo 3D environment. Orientation, projection depth and shadow/lighting effects of each volume render were adjusted for publication quality images. Importantly, these changes were made post-lesion labelling, and did not affect lesion detection or their visual presentation.

We blinded samples at three distinct points in the analysis. First, neonatal CCM model pups were injected with 4-hydroxytamoxifen without knowledge of genotypes. Second, hindbrains from genotyped animals were given randomized labels to provide for blinded microCT scanning by an independent operator. Third, randomized microCT image stacks were analysed in a blinded manner by a single operator not involved in any prior experimental steps.

cDNA constructs and transient expression in HEK293T cells. The human KRIT1-Flag expression plasmid was a gift from D. Marchuk⁸. Human CCM2-Flag, PDGFR10-Flag and MEK3-Flag expression plasmids were purchased from Origene (RC201418, RC200235 and RC210317). Further epitope tag modifications and CCM2 truncations were generated by PCR and conventional cloning methods. The CCM2 CCCTdup mutant was generated by site-directed mutagenesis to generate the exact four-nucleotide duplication (Agilent 200521). One microgram of each construct was transfected in HEK293T cells using Fugene 6 (Promega E2691), and cells were collected 48 h later with gentle lysis buffer (Life Tech 87787) supplemented with protease and phosphatase inhibitors (Roche 11873580001 and 04906845001).

Co-immunoprecipitation and western blotting. Anti-HA antibody (5 µg per co-immunoprecipitation (co-IP), Sigma H3663) conjugated to Protein G Dynabeads (50 µl per co-IP, Life Technologies 10003D) was used for all experiments. Protein (250 µg) from HEK293T cells was used for each co-IP with an incubation period of 30 min at room temperature. HEK293T cells were obtained directly from the ATCC, authenticated and tested negative for mycoplasma. Anti-Flag (1:5,000, Santa Cruz sc-807) and anti-HA (1:5,000, Santa Cruz sc-805) antibodies were used for western blotting and re-probed with anti-GAPDH antibody (1:10,000, Cell Signaling 2118). For co-IP blots of PDGFR10-Flag, horseradish peroxidase (HRP)-conjugated anti-Flag antibody (1:2,000, Sigma A8592) was used to avoid antibody light chain detection.

Expression of tetracycline-inducible CCM2 CCCTdup lentivirus in cultured endothelial cells. Amino-terminal, HA-epitope tagged, human CCM2 CCCTdup was cloned into the pLVX-TRE3G lentiviral vector (Takara Clontech), and lentivirus was generated from co-transfection with packaging plasmids in HEK293T cells as previously described¹². HUVECs were co-infected with CCM2 CCCTdup along with Tet-On 3G lentiviruses and varying doxycycline amounts were added as previously described¹². Cells were collected 48 h after doxycycline addition. Collected cells were pelleted, resuspended, then divided in half. One half was used for total RNA isolated by TRIzol (Life Technologies), and the other half was used for protein extracted by gentle lysis buffer as described above. cDNA was generated from 500 ng total RNA using the Superscript III Reverse Transcriptase (Invitrogen). mRNA levels of the CCM2 CCCTdup lentivirus were assessed by qPCR using both N-terminal and C-terminal human CCM2 qPCR primers as follows: CCM2 N-term forward 5'-CATACCAGGATACCTGAATCCCT-3', CCM2 N-term reverse 5'-AGCTTGACGTTGTACGACAGAC-3'; CCM2 C-term forward 5'-GCCTCTATCCACGAGTTCTGCA-3', CCM2 C-term reverse 5'-AAGTGCTGGCTGTCCTTCTCAG-3'.

Western blotting was performed as previously described with 75 µg protein from HUVEC lysate for each condition.

siRNA knockdown and adenoviral overexpression in cultured endothelial cells. siRNA experiments in HUVECs were performed as previously described¹². HUVECs from Lonza were authenticated and tested negative for mycoplasma. siRNAs directed against *KRIT1* (s2510, Invitrogen), *KLF2* (s20270, Invitrogen), *KLF4* (s17793, Invitrogen), *STK24* (s15993, Invitrogen) or *STK25* (s20570, Invitrogen) were used for the knockdown experiments. For overexpression studies, human microvascular endothelial cells (HMVECs) were infected with adenovirus encoding either mouse *Klf2* (Penn Vector Core), human *KLF4* (Vector BioLabs), or a LacZ control (Penn Vector Core). HMVECs from Lonza were authenticated and tested negative for mycoplasma. Cells were collected 48 h after infection, and total RNA was isolated using TRIzol Reagent (Life Technologies). cDNA was generated from 1 µg total RNA using SuperScript First-strand Synthesis System (Invitrogen) and real-time qPCR was performed using Power SYBR Green PCR Master Mix (Applied Biosciences).

Human primers for qPCR: *KLF2* forward 5'-CTACACCAAGAG TTCGCATCTG-3', *KLF2* reverse 5'-CCGTGTGCTTTCGGTAGTG-3'; *KLF4* forward 5'-AGAGTTCCTCATCAAGGCA-3', *KLF4* reverse 5'-GTCAGTTC ATCTGAGCGGG-3'; *ADAMTS4* forward 5'-CTGACTTCCTGGACAATGGC-3', *ADAMTS4* reverse 5'-GCGGTCAGCATCATAGTCCT-3'.

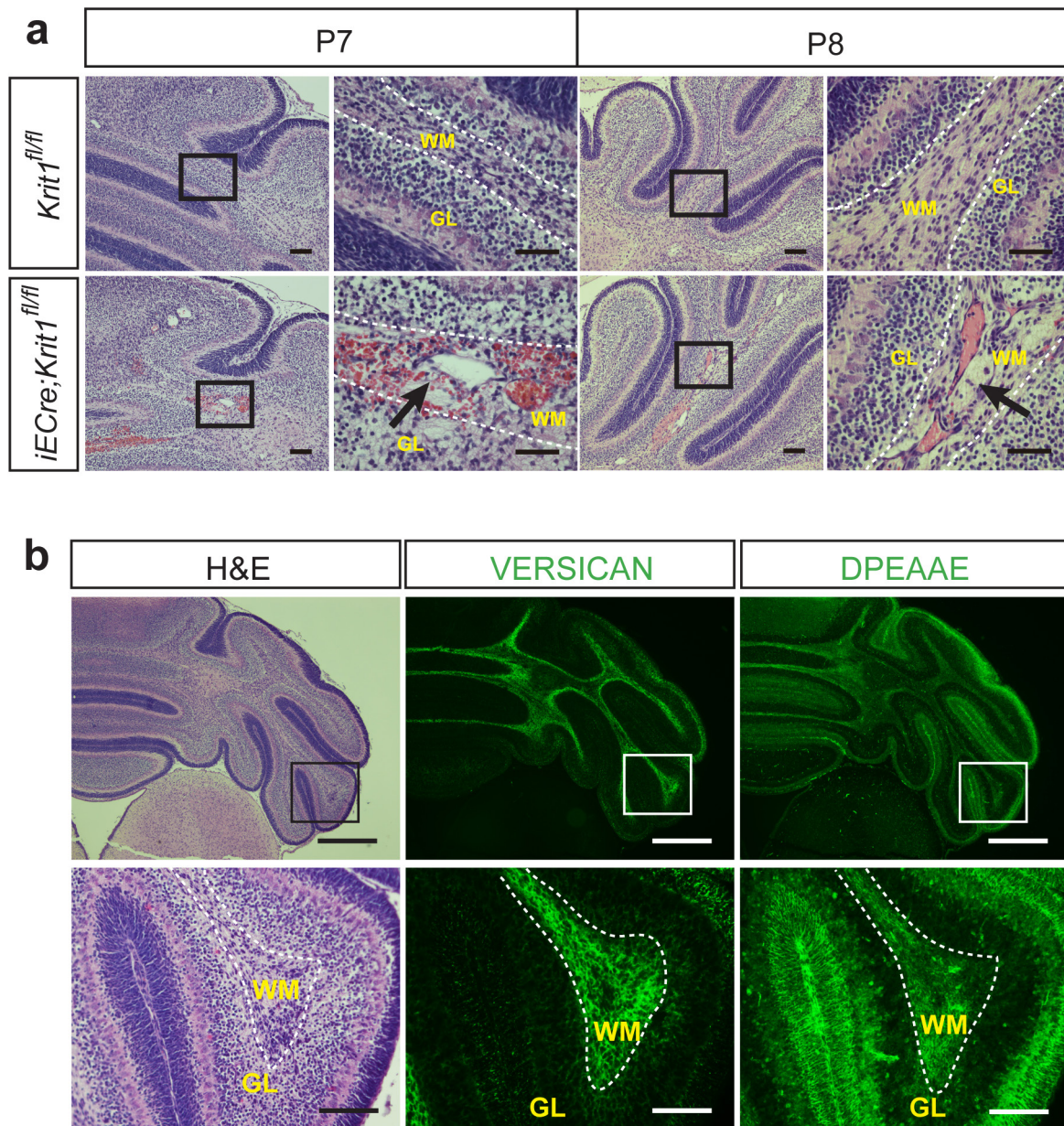
Drug treatments in cultured endothelial cells with loss of KRIT1. After siRNA knockdown of *KRIT1*, HUVECs were treated with BIX02189 (10 µM), XMD17-109 (1 µM), hydroxyfasudil (10 µM), tempol (10 µM) or vitamin D3 (10 µM) for 24 h before collection.

KLF2 and KLF4 immunostaining in human CCM lesions and control brain sections. Sections of human samples were obtained from CCM lesions resected at University of Chicago using Institutional Review Board approved protocols. Control samples were obtained from two autopsy subjects. Rabbit anti-KLF2 (1:200, Abcam ab203591), rabbit anti-KLF4 (1:50, Cell Signaling 4038S) and

mouse anti-PECAM-1 (1:1,000, Cell Signaling 3528S) were used for immunostaining.

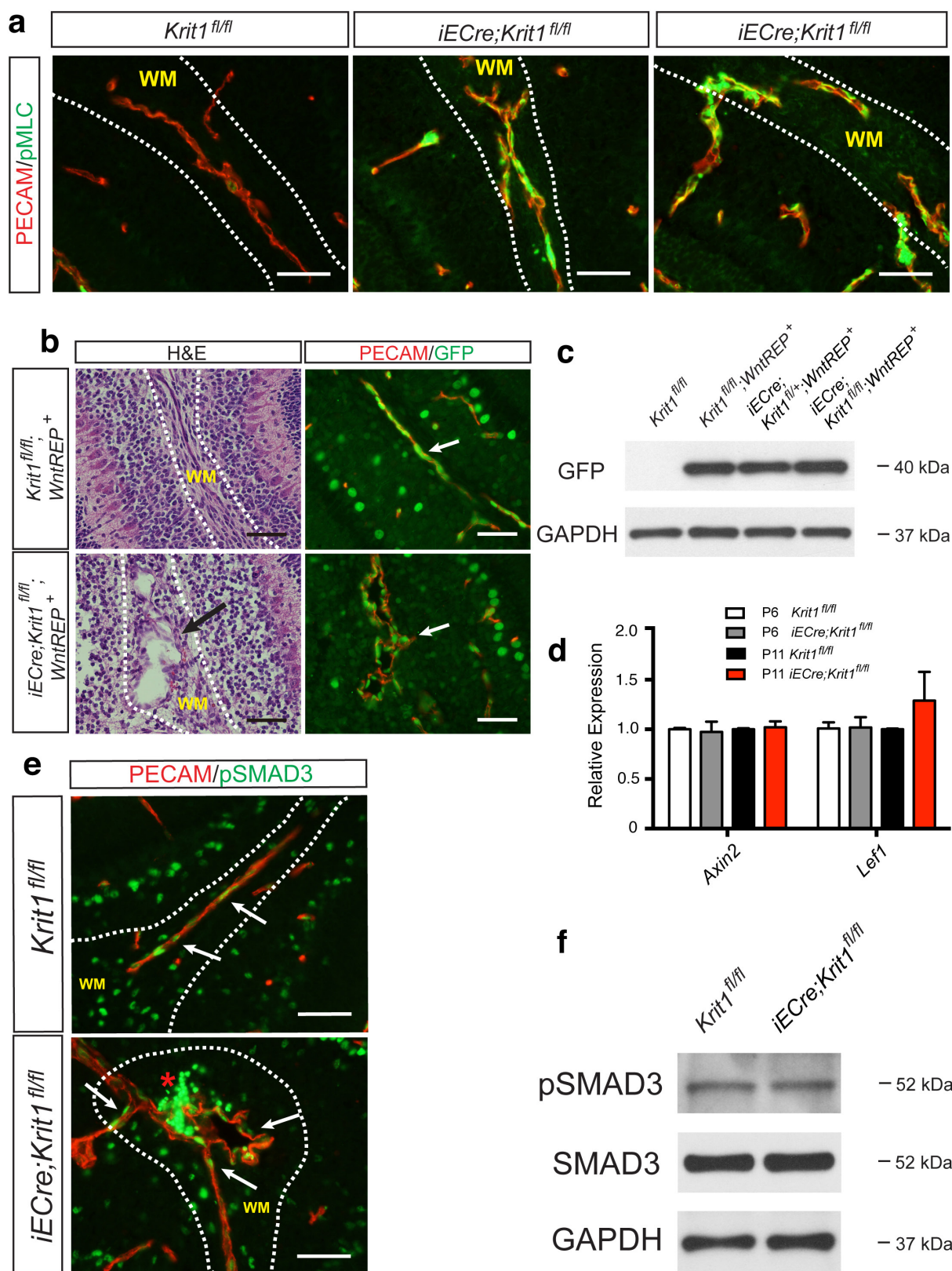
Patient information pertaining to the human CCM samples studied is summarized below. KRIT1 familial: female, age 11, with the CCM1 Common Hispanic Mutation, removed owing to CCM haemorrhage. PDCD10 familial: female, age 19, *PDCD10* 474+5G>A mutation, removed owing to CCM haemorrhage. Sporadic 1: female, age 24, solitary sporadic lesion, removed owing to headaches and lesion growth. Sporadic 2: male, age 31, solitary sporadic lesion, removed owing to seizures. **Statistics.** Sample sizes were estimated based on our preliminary findings for rescue of the CCM phenotype. These were estimated from visual analysis of the hindbrains of P11 animals from which lesion numbers and sizes can be directly assessed, an observation corroborated by microCT-based volume rendering of lesions. These preliminary assessments suggested that there were very large differences between genetic rescue and non-rescue samples (conservatively 80% rescue) that would allow statistical interpretation with relatively small *n* values. All experiments and controls were littermates, and none was excluded from analysis. *P* values were calculated using an unpaired two-tailed Student's *t*-test, analysis of variance (ANOVA), or chi-square analysis as indicated. The mean and s.e.m. are shown in the bar graphs.

30. Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465**, 483–486 (2010).
31. Mleynek, T. M. *et al.* Lack of CCM1 induces hypersprouting and impairs response to flow. *Hum. Mol. Genet.* **23**, 6223–6234 (2014).
32. Lee, J. S. *et al.* Klf2 is an essential regulator of vascular hemodynamic forces *in vivo*. *Dev. Cell* **11**, 845–857 (2006).
33. Katz, J. P. *et al.* The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development* **129**, 2619–2628 (2002).
34. Kisanuki, Y. Y. *et al.* Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis *in vivo*. *Dev. Biol.* **230**, 230–242 (2001).



Extended Data Figure 1 | Histological characteristics of early CCM lesions and the cerebellar white matter in which they form. a, P7 and P8 CCM lesions in the *Krit1* model. Hindbrains from *Krit1^{ECKO}* mice were collected at P7 and P8, and sections were stained with haematoxylin and eosin to detect vascular malformations. Images are representative of 4 independent experiments. Dotted lines indicate the cerebellar white matter; arrows indicate CCM lesions. **b,** Versican is abundant in the white matter of the P7 hindbrain, the site of primary CCM formation in the neonatal mouse model. Haematoxylin and eosin staining of the hindbrain at low (top left) and high (bottom left) magnifications. Antibody staining

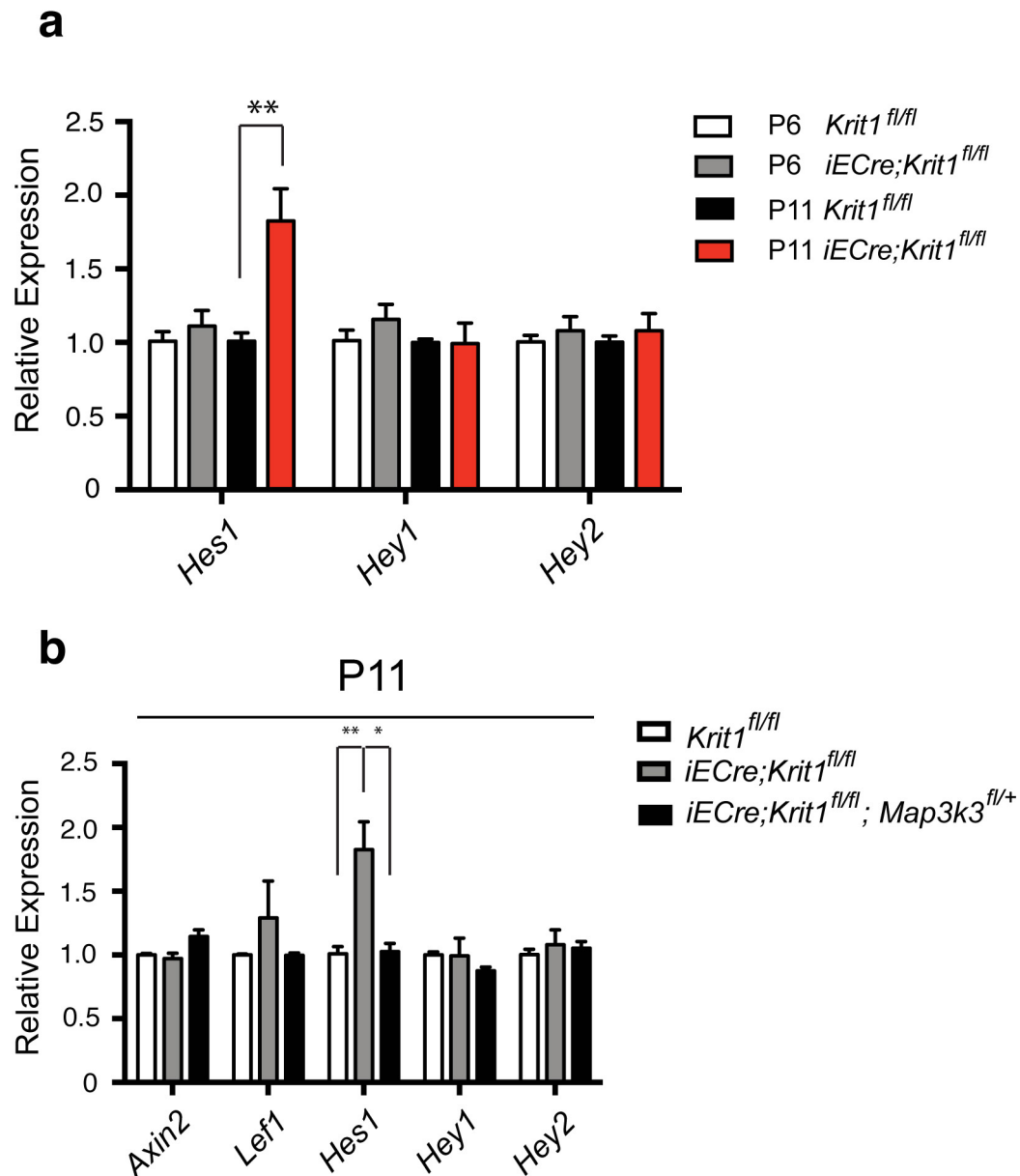
reveals abundant versican protein in the white matter of the hindbrain (centre), the site at which CCM lesions specifically appear at this time point in the *Krit1* mouse model, and less abundant versican in the cortex. DPEAAE staining reveals a pattern of ADAMTS-mediated versican degradation that is the inverse of intact versican; that is, higher in the cortex and lower in the white matter (right). Results are representative of 3 independent experiments. Scale bars, 500 μ m (top) and 100 μ m (bottom). Boxes indicate regions shown at higher magnification below; dotted lines indicate the cerebellar white matter.



Extended Data Figure 2 | Endothelial Rho activity, but not β -catenin or SMAD3 signalling, is increased during CCM formation.

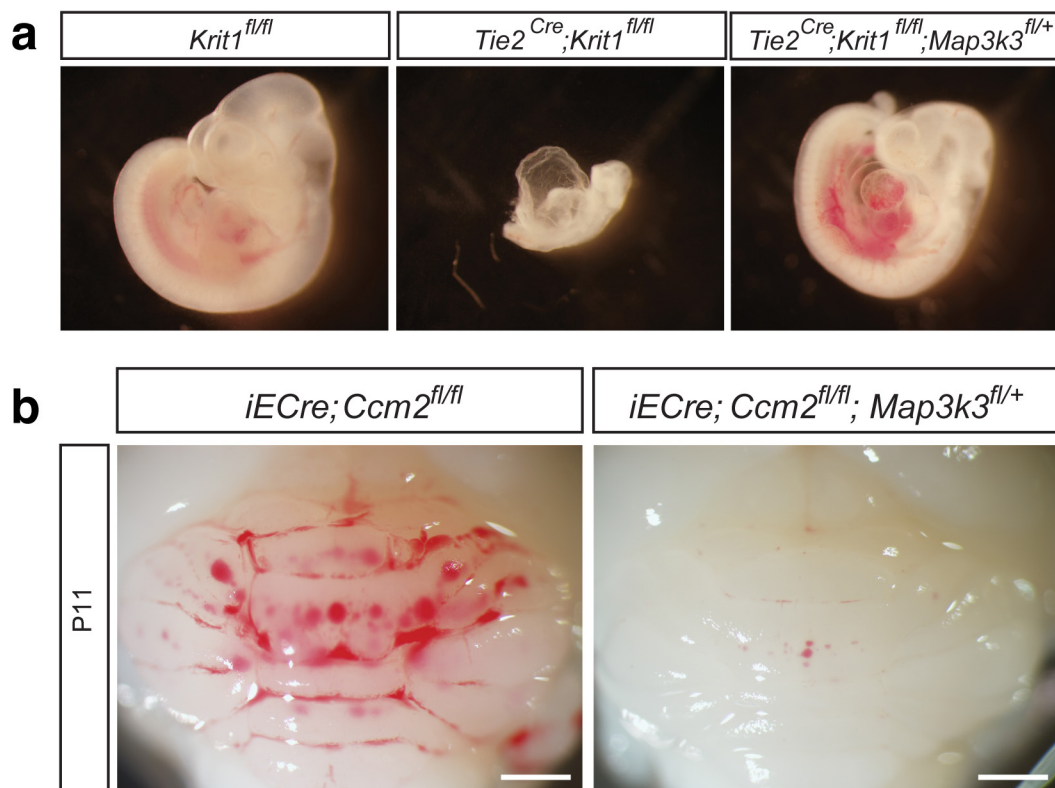
a, Immunostaining for the endothelial marker PECAM and pMLC in the white matter of the cerebellum of P6 control and *Krit1^{ECKO}* littermates. Scale bars, 50 μ m. **b**, Anti-GFP immunostaining to detect TCF/Lef:H2B-GFP Wnt/ β -catenin reporter (WntREP) activity. Scale bars, 50 μ m. **c**, Immunoblotting of P6 brain endothelial cell lysate for GFP. Results

are representative of 3 independent experiments. **d**, qPCR analysis of β -catenin target genes in hindbrain endothelial cells ($n = 4-5$). $P > 0.05$ (unpaired two-tailed Student's t -test) for comparison of all values. Error bars indicate s.e.m. **e**, Immunostaining for phosphorylated-SMAD3 (pSMAD3) and PECAM. Scale bars, 50 μ m. **f**, Total SMAD3 and pSMAD3 were detected by immunoblotting cerebellar endothelial cell lysates. Results are representative of 3 independent experiments.



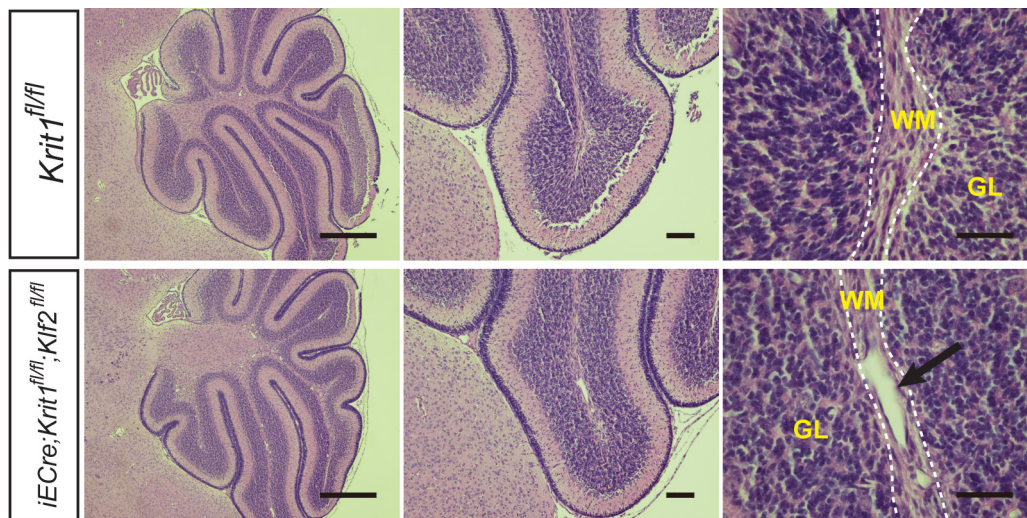
Extended Data Figure 3 | *Hes1* expression rises after primary CCM lesion formation in *Krit1*^{ECKO} hindbrain endothelial cells, and is reversed by *Map3k3* haploinsufficiency. a, qPCR analysis of Notch target genes in endothelial cells freshly collected from the hindbrain of P6 or P11 *Krit1*^{ECKO} animals compared with control, tamoxifen-treated littermates

($n = 4$). $**P < 0.01$ (unpaired two-tailed Student's t -test). **b,** qPCR analysis of gene expression in endothelial cells freshly collected from the cerebellum of P11 *Krit1*^{fl/fl} and *Krit1*^{ECKO} animals, with and without endothelial loss of one *Map3k3* allele ($n = 4$). $**P < 0.01$ (unpaired two-tailed Student's t -test). Error bars indicate s.e.m.



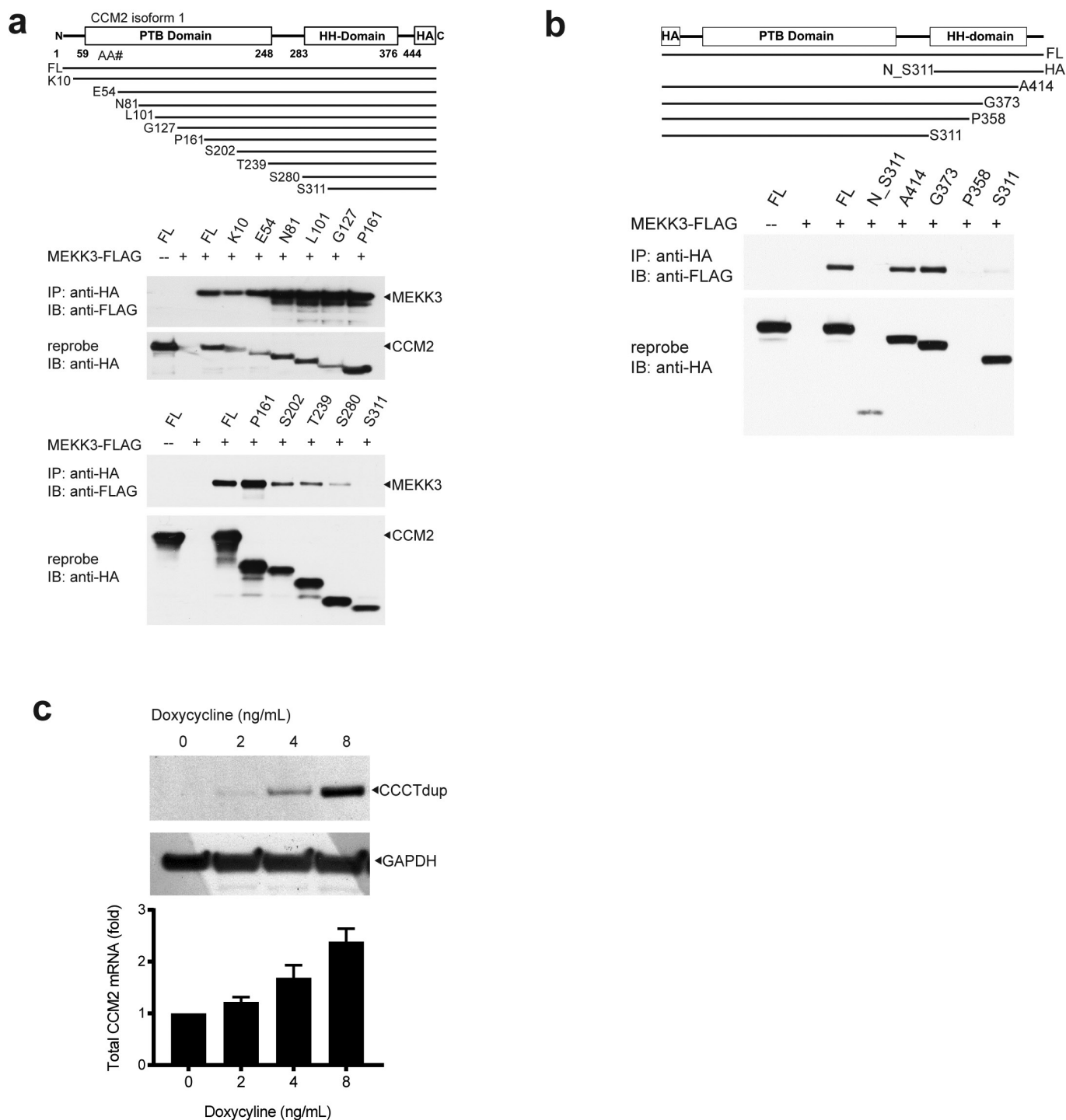
Extended Data Figure 4 | *Map3k3* haploinsufficiency rescues both early embryonic lethality due to endothelial loss of *Krit1* and postnatal CCM lesion formation due to endothelial loss of *Ccm2*. **a**, Partial loss of MEKK3 rescues early lethality associated with endothelial deletion of *Krit1*. E10.5 littermate embryos are shown at the same magnification. *Tie2*^{Cre};*Krit1*^{fl/fl} animals lacking endothelial KRIT1 die by E9.5 due to lack of patent branchial arch arteries (middle). By contrast, *Tie2*^{Cre};*Krit1*^{fl/fl};*Map3k3*^{fl/+} animals develop patent arteries and survive past

mid-gestation (right). Note the presence of pericardial oedema due to a persistent cardiac defect in the *Tie2*^{Cre};*Krit1*^{fl/fl};*Map3k3*^{fl/+} embryo. Images are representative of 3 independent experiments. **b**, Rescue of CCM formation in the *Ccm2* model with loss of MEKK3. *iECre*;*Ccm2*^{fl/fl};*Map3k3*^{+/+} (*Ccm2*^{ECKO}) and *iECre*;*Ccm2*^{fl/fl};*Map3k3*^{fl/+} neonates were induced with tamoxifen at P1 and lesion formation scored by visual detection in the hindbrain at P11 as done for *Krit1*^{ECKO} animals. Images are representative of 4 independent experiments. Scale bars, 1 mm.



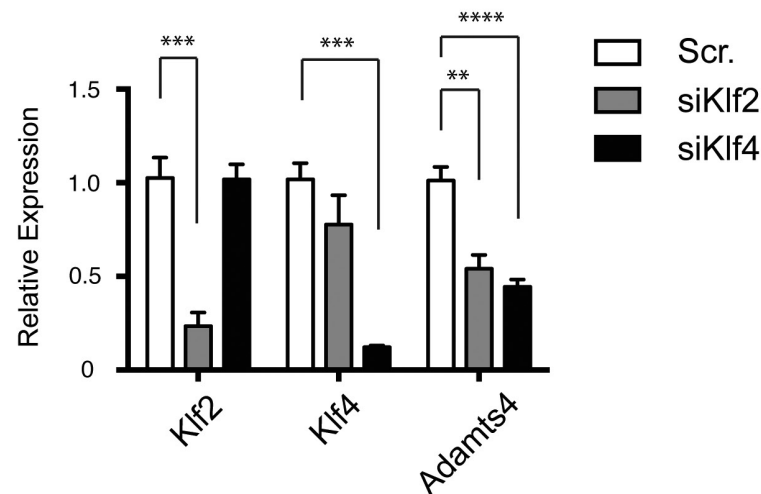
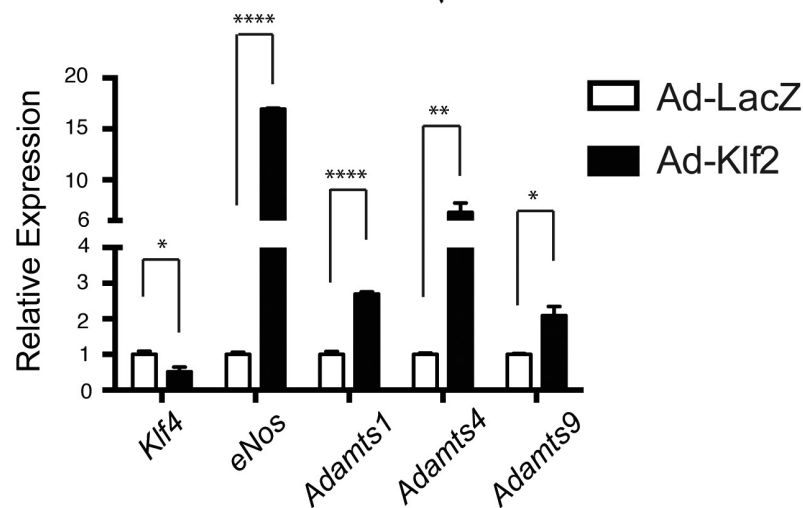
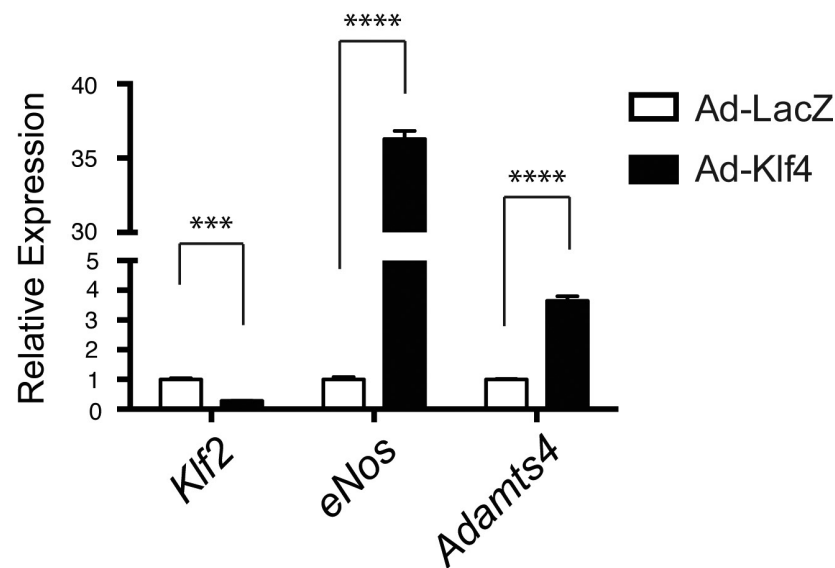
Extended Data Figure 5 | Rescue of CCM formation in *Krt1^{ECKO}* animals with homozygous loss of *Klf2*. Haematoxylin and eosin sections through the hindbrain of P11 *iECre;Krt1^{fl/fl};Klf2^{fl/fl}* and control *iECre;Krt1^{fl/fl}* animals are shown. No true CCM lesions were observed

but venous dilatation was evident (arrow). Images are representative of 3 independent experiments. Dotted lines indicate the cerebellar white matter. Scale bars, 500 μm (left), 100 μm (middle) and 50 μm (right).



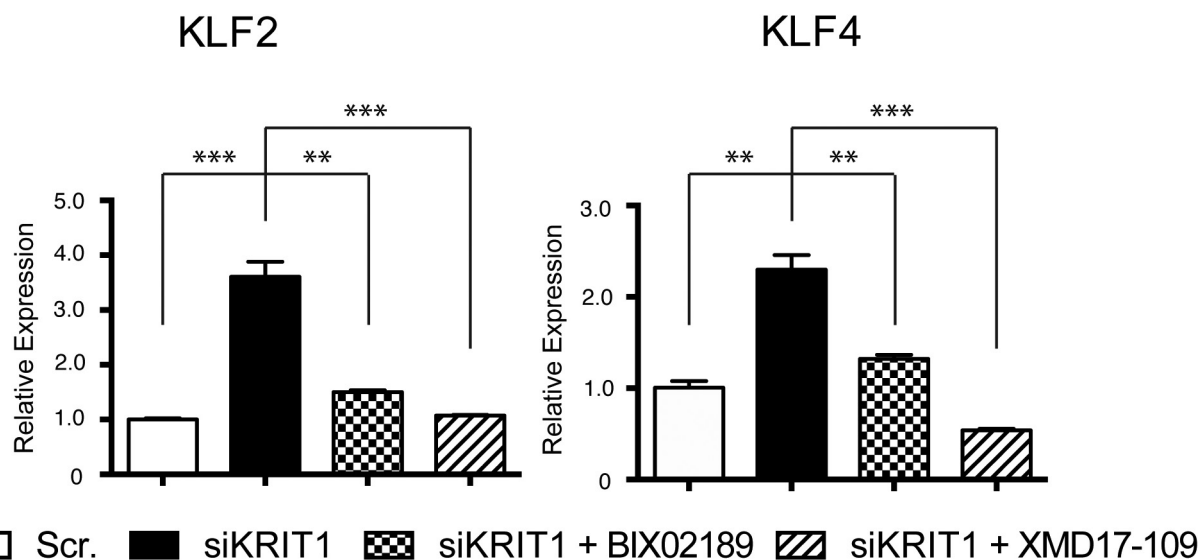
Extended Data Figure 6 | CCM2 binds MEKK3 via its C-terminal helical harmonin domain. **a**, MEKK3 binds the C-terminal domain of CCM2. Top, schematic of the CCM2 protein domains. Bottom, series of N-terminal-truncated CCM2-HA proteins used to map the CCM2-MEKK3 binding region. MEKK3-Flag and the indicated N-terminally truncated CCM2-HA proteins were co-expressed in HEK293T cells before immunoprecipitation with anti-HA antibodies and immunoblotting with anti-Flag antibodies to detect associated MEKK3 proteins (top). Immunoprecipitated HA-CCM2 was detected with anti-HA antibodies (below). **b**, MEKK3-Flag and the indicated C-terminally truncated

CCM2-HA proteins were co-immunoprecipitated to map the CCM2-MEKK3 binding region. **c**, CCM2 CCCTdup protein is stably expressed in cultured endothelial cells. CCM2 CCCTdup-HA protein was expressed in cultured HUVECs using a tetracycline-inducible lentiviral vector at varying doses of doxycycline. The total (that is, endogenous CCM2 plus viral CCM2 CCCTdup) CCM2 mRNA levels were measured using qPCR (bottom), and CCM2 CCCTdup protein detected using anti-HA immunoblotting (top). GAPDH immunoblots are shown for loading control.

a**b****c**

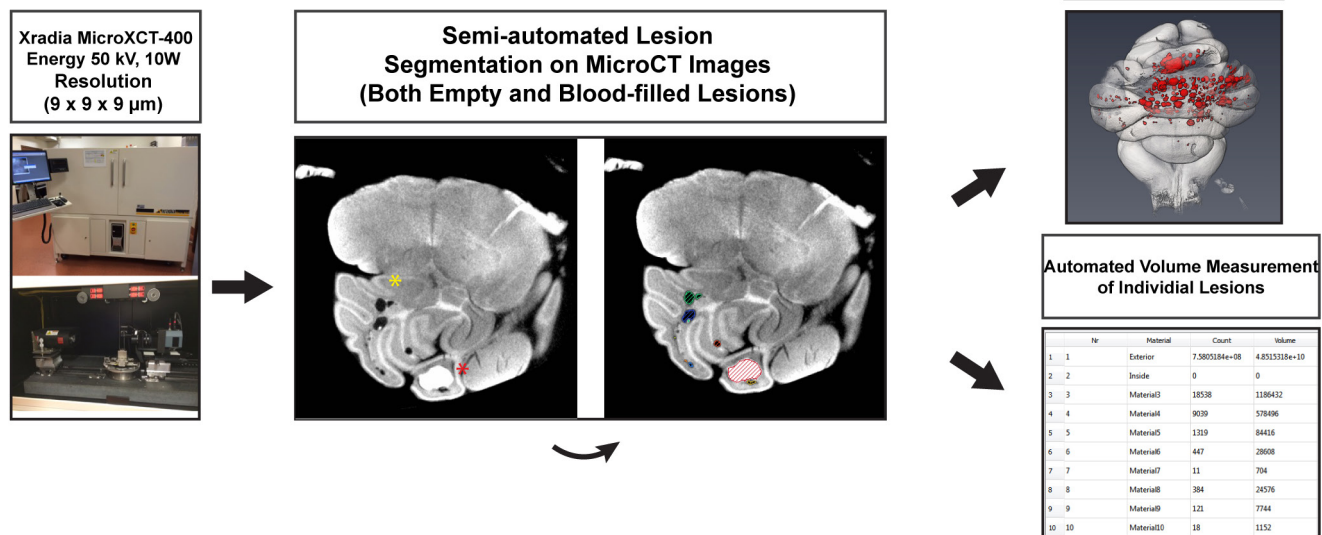
Extended Data Figure 7 | The versican-degrading ADAMTS proteases are regulated by KLF2 and KLF4. **a**, siRNA knockdown of *KLF2* or *KLF4* in cultured HUVECs reduces *ADAMTS4* expression, as measured by qPCR. **b**, Adenoviral expression of KLF2 drives expression of the versican-degrading proteases *ADAMTS1*, *ADAMTS4* and *ADAMTS9* in addition to the known KLF2/4 target gene *eNOS* (also known as *NOS3*). *KLF4* levels are reduced in KLF2-overexpressing HUVECs. Results are shown relative

to expression following exposure to control adeno-LacZ virus. **c**, Adenoviral expression of KLF4 drives expression of *ADAMTS4* in addition to *eNOS*. *KLF2* levels are reduced in KLF4-overexpressing HUVECs. Results are shown relative to expression after exposure to control adeno-LacZ virus ($n = 4-5$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (unpaired two-tailed Student's *t*-test). Error bars indicate s.e.m.



Extended Data Figure 8 | The MEK5 inhibitor BIX02189 and the ERK5 inhibitor XMD17-109 reverse the increase in *KLF2* and *KLF4* expression conferred by loss of KRIT1 in cultured endothelial cells. HUVECs were treated with scrambled or *KRIT1*-targeted siRNAs with

or without BIX02189 (10 μ M) and XMD17-109 (1 μ M) (MEK5 and ERK5 inhibitors, respectively) for 24 h, and the levels of *KLF2* and *KLF4* mRNA were measured by qPCR ($n = 3$). ** $P < 0.01$; *** $P < 0.001$ (unpaired two-tailed Student's t -test). Error bars indicate s.e.m.



Extended Data Figure 9 | Imaging and volume measurement of mouse hindbrain CCM lesions using microCT. The imaging of P11 mouse hindbrains and analysis of the raw data to create composite CCM lesion images and volumes are shown.

Structural basis of lenalidomide-induced CK1 α degradation by the CRL4^{CRBN} ubiquitin ligase

Georg Petzold^{1,2}, Eric S. Fischer^{1,2†} & Nicolas H. Thomä^{1,2}

Thalidomide and its derivatives, lenalidomide and pomalidomide, are immune modulatory drugs (IMiDs) used in the treatment of haematologic malignancies^{1,2}. IMiDs bind CRBN, the substrate receptor of the CUL4–RBX1–DDB1–CRBN (also known as CRL4^{CRBN}) E3 ubiquitin ligase³, and inhibit ubiquitination of endogenous CRL4^{CRBN} substrates⁴. Unexpectedly, IMiDs also repurpose the ligase to target new proteins for degradation. Lenalidomide induces degradation of the lymphoid transcription factors Ikaros and Aiolos (also known as IKZF1 and IKZF3)^{5–7}, and casein kinase 1 α (CK1 α)⁸, which contributes to its clinical efficacy in the treatment of multiple myeloma⁵ and 5q-deletion associated myelodysplastic syndrome (del(5q) MDS)⁸, respectively. How lenalidomide alters the specificity of the ligase to degrade these proteins remains elusive. Here we present the 2.45 Å crystal structure of DDB1–CRBN bound to lenalidomide and CK1 α . CRBN and lenalidomide jointly provide the binding interface for a CK1 α β -hairpin-loop located in the kinase N-lobe. We show that CK1 α binding to CRL4^{CRBN} is strictly dependent on the presence of an IMiD. Binding of IKZF1 to CRBN similarly requires the compound and both, IKZF1 and CK1 α , use a related binding mode. Our study provides a mechanistic explanation for the selective efficacy of lenalidomide in del(5q) MDS therapy^{8,9}. We anticipate that high-affinity protein–protein interactions induced by small molecules will provide opportunities for drug development, particularly for targeted protein degradation.

Thalidomide, lenalidomide and pomalidomide are used in the treatment of multiple myeloma^{1,2}. In del(5q) MDS patients, lenalidomide shows pronounced clinical efficacy compared to thalidomide and pomalidomide^{2,9,10}, which has been linked to induced degradation of CK1 α (ref. 8). Del(5q) MDS is a haematologic disorder in which deletion of the long arm of chromosome 5 eliminates one casein kinase 1 alpha allele (CSNK1A1). Heterozygous deletion of CSNK1A1 causes haematopoietic stem cell hyperproliferation, whereas homozygous loss of both alleles results in stem cell failure and apoptosis¹¹. CSNK1A1 haploinsufficiency thus sensitizes malignant del(5q) MDS cells to lenalidomide-induced CK1 α degradation, while wild-type cells remain unaffected⁸.

To characterize the binding of CK1 α to CRL4^{CRBN} *in vitro*, we reconstituted the complex and measured the affinity of neddylated CRL4^{CRBN} (N_8 CRL4^{CRBN}) to CK1 α using time-resolved fluorescence resonance energy transfer (TR-FRET) (Extended Data Fig. 1a, b; see Methods). At saturating lenalidomide concentrations (Extended Data Fig. 1c), N_8 CRL4^{CRBN} and CK1 α tightly associate with an apparent dissociation constant (K_D^{app}) of 242 ± 15 nM (Fig. 1a and Extended Data Fig. 1d). In contrast, N_8 CRL4^{CRBN} did not bind CK1 α in the absence of lenalidomide, or when rendered IMiD-binding deficient (IBD) by mutations in the IMiD-binding domain of CRBN (Fig. 1a and Extended Data Fig. 1e). In the presence of IMiD analogues, pomalidomide and thalidomide, N_8 CRL4^{CRBN} binding to CK1 α was weakened

~ 2 -fold (K_D^{app} (pomalidomide) = 608 ± 60 nM) and ~ 5 -fold (K_D^{app} (thalidomide) = $1,122 \pm 109$ nM) compared to lenalidomide (Fig. 1b and Extended Data Fig. 1c).

To understand the molecular basis of CK1 α degradation in del(5q) MDS, we pursued the high-affinity complex of CRL4^{CRBN}, CK1 α and lenalidomide for structural studies. The modular CRL4^{CRBN} assembly adopts a U-shaped architecture⁴, in which the DDB1 adaptor protein mediates interactions between CRBN and the CUL4–RBX1 ligase

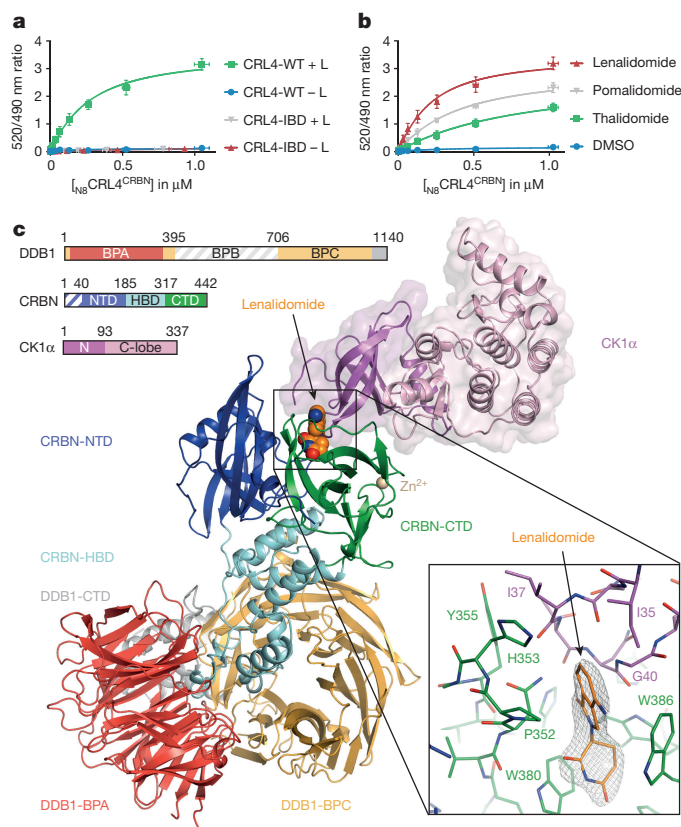


Figure 1 | Overall structure of the DDB1 Δ BPP–CRBN–lenalidomide–CK1 α complex. a, TR-FRET. Wild-type (WT) or IMiD binding deficient (IBD) mutant N_8 CRL4^{CRBN} titrated to CK1 α in the absence or presence of 5 μ M lenalidomide (L). b, TR-FRET. N_8 CRL4^{CRBN}–CK1 α complex formation in the absence (DMSO) or presence of 5 μ M lenalidomide, pomalidomide or thalidomide. TR-FRET data in this figure are biological replicates presented as means \pm s.d. ($n = 3$). c, Cartoon representation of DDB1 Δ BPP–CRBN–lenalidomide–CK1 α . Top left, DDB1, CRBN and CK1 α domain colouring and boundaries. Hatched lines indicate sequences omitted in the crystallized proteins. Bottom right, close-up of the CRBN–lenalidomide–CK1 α interface. Composite omit $2F_o - F_c$ map contoured at 2.0σ .

¹Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. ²University of Basel, Petersplatz 10, CH-4003 Basel, Switzerland. [†]Present addresses: Department of Cancer Biology, Dana-Farber Cancer Institute, LC-4312, 360 Longwood Avenue, Boston, Massachusetts 02215, USA; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02215, USA.

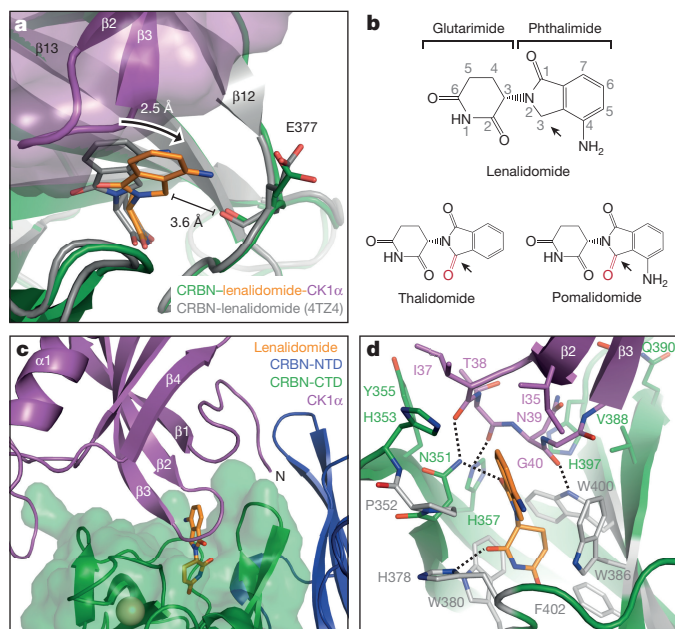


Figure 2 | CK1 α interacts with CRBN and lenalidomide.

a, Superposition of CRBN–lenalidomide–CK1 α and human CRBN bound to lenalidomide (PDB entry 4TZ4). **b**, Chemical structures of lenalidomide, pomalidomide and thalidomide. Arrows indicate C3 position of the phthalimide ring. **c**, CRBN binds the glutarimide moiety in a hydrophobic pocket. The phthalimide ring of lenalidomide is surface exposed. A β -hairpin loop of CK1 α binds CRBN on top of its lenalidomide-binding pocket. **d**, Side-chain interactions between CK1 α , CRBN and lenalidomide. Lenalidomide provides van der Waals interactions with CK1 α . CRBN residues that form the hydrophobic pocket are shown in grey. Dashed lines indicate hydrogen bonds.

arm. This arrangement juxtaposes the CRBN substrate receptor to RBX1, the subunit that recruits the activated E2–Ub to the complex. DDB1 is composed of three β -propeller domains (BPA, BPB, BPC). Only two β -propellers (BPA, BPC) are necessary and sufficient for CRBN binding^{4,12}. The BPB domain interacts with CUL4 and was removed to facilitate crystallization (DDB1 Δ BPB). Crystals were obtained for the 178-kDa DDB1 Δ BPB–CRBN–lenalidomide–CK1 α complex and its structure was determined to 2.45 Å resolution (Extended Data Table 1). The unit cell of the crystal contained two identical complexes, one of which refined to considerably lower B factors (chain B and chain C; Extended Data Fig. 2a–c).

We find that CK1 α exclusively interacts with the C-terminal domain of CRBN (CTD; residues 317–442), which accommodates lenalidomide in a hydrophobic pocket⁴ (Fig. 1c). CRBN and lenalidomide together provide the interaction surface for CK1 α (CRBN–CK1 α , 718.1 Å²; CRBN–lenalidomide, 236.6 Å²; lenalidomide–CK1 α , 96.8 Å²; total, 1,051.5 Å²). No contacts are observed between CK1 α and DDB1. The kinase domain of CK1 α adopts an active conformation¹³, although the nucleotide binding pocket is not occupied by ATP (Extended Data Fig. 3). CK1 α binding repositions the phthalimide ring of lenalidomide 2.5 Å towards CRBN residue Glu377, which resides in the loop linking β -stands 12 and 13 (Fig. 2a). This shifts the lenalidomide C3 phthalimide to the backbone carbonyl of Glu377. Thalidomide and pomalidomide carry a carbonyl group at their C3 phthalimide position (Fig. 2b) that would result in clashes with the CRBN backbone (Extended Data Fig. 4). This observation provides an unanticipated rationale for the preference of CK1 α for lenalidomide during complex formation with CRL4^{CRBN} (Fig. 1b), and explains how a buried carbonyl, absent in lenalidomide and distant from the IMiD–CK1 α interface, contributes to target specificity.

CK1 α binds CRBN–lenalidomide via a β -hairpin loop (residues 35–41) in the kinase N-lobe that connects the CK1 α β -strands 2 and 3

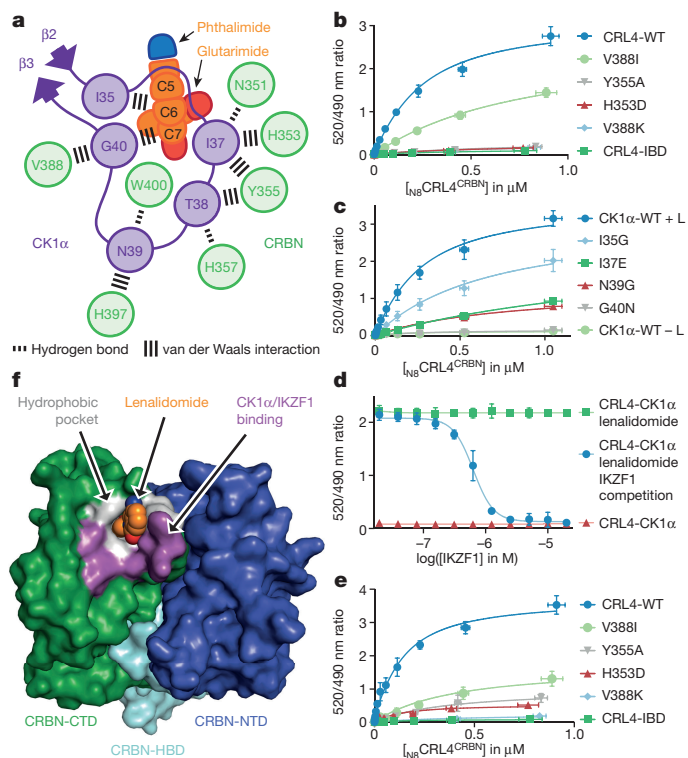


Figure 3 | CK1 α and IKZF1 share a common CRBN interface.

a, Schematic CRBN–lenalidomide–CK1 α interface. **b**, Wild-type or mutant N₈CRL4^{CRBN} titrated to CK1 α in the presence of lenalidomide. **c**, N₈CRL4^{CRBN} titrated to mutant CK1 α in presence of lenalidomide (L). **d**, Unlabelled IKZF1^{83–196/239–255} (0.02–20 μ M) titrated to pre-assembled N₈CRL4^{CRBN}–lenalidomide–CK1 α complexes (500 nM N₈CRL4^{CRBN}, 100 nM CK1 α , 5 μ M lenalidomide; blue graph). N₈CRL4^{CRBN}–lenalidomide–CK1 α (green). N₈CRL4^{CRBN}, CK1 α without lenalidomide (red). **e**, Wild-type or mutant N₈CRL4^{CRBN} titrated to wild-type IKZF1 in the presence of pomalidomide. TR-FRET data (b–e) are biological replicates presented as means \pm s.d. (n = 3). **f**, CRBN surface contacted by CK1 α and IKZF1.

(Fig. 2c). Residues within this loop interact with the otherwise solvent-exposed phthalimide ring of lenalidomide and contact CRBN residues that surround the lenalidomide binding pocket (Fig. 2d and Fig. 3a). Mutation of CRBN residues His353 to aspartate and Tyr355 to alanine did not prevent IMiD binding (Extended Data Fig. 1e), but affected CK1 α complex formation (Fig. 3b and Extended Data Fig. 4f). The side chain of CK1 α Ile35 provides hydrophobic contact with C5 and C6 of the lenalidomide phthalimide ring and CK1 α Gly40 is sandwiched between phthalimide C7 and the side chain of CRBN Val388 (Figs 2d and 3a). Sequence alignments between casein kinase 1 and 2 isoforms reveal substitution of Gly40 for a bulky asparagine residue in CK2 α (Extended Data Fig. 5a, b), and mutation of CK1 α Gly40 to asparagine abolished CRBN binding (Fig. 3c). Consistently, levels of CK2 α are not reduced in cells treated with lenalidomide (Extended Data Fig. 5c). Together, these observations emphasize the importance of the β -hairpin residue Gly40 for CK1 α degradation.

The side chain of CRBN Val388 is in close proximity to the CK1 α residues Gly40 and Glu41. Val388 is substituted for isoleucine in CRBN from chicken (Ile390) and mouse (Ile391; Extended Data Fig. 5d, e). Mutation of Val388 to isoleucine in human CRBN led to an \sim 4-fold reduction of the apparent K_D (1,095 \pm 99 nM; Fig. 3b), which is in agreement with steric clashes expected upon substitution (Extended Data Fig. 5e). Similarly, the Val388Lys mutation rendered CRBN unable to interact with CK1 α *in vitro* (Fig. 3b). Mouse cells with the naturally occurring CRBN Ile391 are largely resistant to IMiD-induced CK1 α degradation⁸. The Ile391Val mutation in mouse CRBN alone was sufficient to restore IMiD-induced degradation of CK1 α and

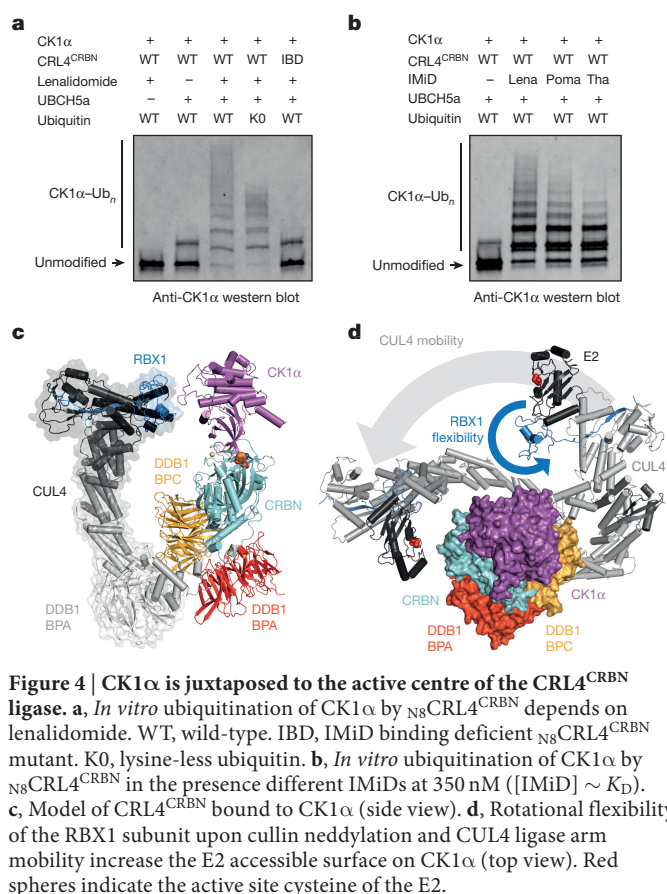


Figure 4 | CK1α is juxtaposed to the active centre of the CRL4^{CRBN} ligase. **a**, *In vitro* ubiquitination of CK1α by N₈CRL4^{CRBN} depends on lenalidomide. WT, wild-type. IBD, IMiD binding deficient N₈CRL4^{CRBN} mutant. K0, lysine-less ubiquitin. **b**, *In vitro* ubiquitination of CK1α by N₈CRL4^{CRBN} in the presence different IMiDs at 350 nM ([IMiD] ~ K_D). **c**, Model of CRL4^{CRBN} bound to CK1α (side view). **d**, Rotational flexibility of the RBX1 subunit upon cullin neddylation and CUL4 ligase arm mobility increase the E2 accessible surface on CK1α (top view). Red spheres indicate the active site cysteine of the E2.

that of two additional CRL4^{CRBN} substrates, IKZF1 and IKZF3 (ref. 8), which is consistent with our structural and biochemical findings.

Next we examined the spatial relationship between the binding sites of CK1α and the Ikaros family transcription factors on CRBN. A truncated form of human IKZF1 (see Methods) bound N₈CRL4^{CRBN} with a higher affinity (K_D^{app} (lenalidomide) 101 ± 2 nM) than CK1α (Extended Data Fig. 6) and efficiently displaced CK1α from CRBN (IC_{50} , 712 ± 11 nM; K_i , 89 ± 3 nM; Fig. 3d). Thus, CK1α and IKZF1 recruitment is mutually exclusive. We then tested the effects of different IMiDs on IKZF1 recruitment and observed that lenalidomide ($K_D^{app} = 101 \pm 2$ nM) and pomalidomide ($K_D^{app} = 96 \pm 24$ nM) mediated IKZF1 binding to CRBN with comparable affinities, whereas in the presence of thalidomide the apparent K_D was reduced ~ 1.5 -fold ($K_D^{app} = 155 \pm 33$ nM) (Extended Data Fig. 6a). No binding between IKZF1 and CRBN was detected in the absence of IMiDs. Recruitment of CK1α and IKZF1 to CRL4^{CRBN} thus strictly depends on the presence of an IMiD, and occurs through overlapping binding sites on CRBN (Fig. 3e, f).

IKZF1 contains four amino-terminal C2H2-type zinc-fingers¹⁴ (residues 118–225; Extended Data Fig. 7a). IKZF1 Gln146 within zinc-finger 2 (ZF2), when replaced by histidine, was shown to confer resistance to IMiD-induced degradation in cells⁵. IKZF1 ZF2 has strong sequence homology to a known zinc-finger structure (PDB 2113) (Extended Data Fig. 7b). The β -hairpin loop (residues 146–151) in the predicted structure of IKZF1 ZF2 shows striking conformational similarity to that of CK1α (residues 35–40; Extended Data Fig. 7c). IKZF1 residue Gln146 is part of this loop, and the Gln146His mutation impaired binding to CRBN–pomalidomide *in vitro* (Extended Data Fig. 7d). There is little sequence conservation between CK1α and IKZF1 with the exception of Gly151 (equivalent to Gly40 in CK1α; Extended Data Fig. 7b). Mutation of IKZF1 Gly151 (Gly151Asn) or two zinc-coordinating residues (Cys147Ala, His167Ala) in ZF2 ablate pomalidomide-dependent IKZF1 binding to CRBN, as well as

substitution of Leu166 (Leu166Glu; Extended Data Fig. 7d). The additional contact mediated by Leu166 seems unique to IKZF1 (Extended Data Fig. 7e–i). The overlay of CK1α with IKZF1 ZF2 further predicts that DNA bound by IKZF1 would clash with CRBN (Extended Data Fig. 8a). In electrophoretic mobility shift assays and TR-FRET experiments, we observed that recruitment to DDB1–CRBN–pomalidomide is impaired when IKZF1 is engaged with its consensus DNA (Extended Data Fig. 8b–d). Thus, our results strongly suggest that IKZF1 uses a structurally similar β -hairpin loop containing the glycine signature for binding the CRBN–IMiD interface, which explains mutual exclusive binding of IKZF1 and CK1α. Additional zinc-finger proteins were found downregulated in the presence of lenalidomide (for example ZNF692, ZFP91) (ref. 8), and these carry similar structural motifs.

Lenalidomide recruits CK1α to CRL4^{CRBN}, which results in CK1α degradation in MDS-L cells⁸. Experiments in a fully reconstituted system further demonstrate that CK1α binding to N₈CRL4^{CRBN} is sufficient to drive its ubiquitination *in vitro* (Fig. 4a). In case of lenalidomide, high-affinity CK1α binding leads to polyubiquitination (Fig. 4b), whereas weaker binding in the presence of pomalidomide and thalidomide (Fig. 1b and Fig. 2a, b) results in reduction of higher-order ubiquitin conjugates. This difference is only evident at non-saturating IMiD concentrations (Fig. 4b and Extended Data Fig. 9a). We expect this to be the case also in cells owing to the reported low solubility and permeability of IMiDs^{15–18}. Assuming comparable concentrations of the three IMiDs in cells, differences in CK1α binding coupled to processive ubiquitination may account for the strong cellular preference for lenalidomide over pomalidomide and thalidomide in targeting CK1α for degradation⁸.

Given the modular nature of CRL4 ligases^{19,20}, the DDB1 Δ BPB–CRBN–lenalidomide–CK1α complex can be combined with existing DDB1–CUL4–RBX1 structures into a model of the entire CRL4^{CRBN}–lenalidomide–CK1α assembly (Fig. 4c). The unique flexibility of the CUL4 ligase arm combined with unhooking of RBX1 upon cullin neddylation²¹ creates an exceptionally large ubiquitination zone¹⁹ ($340 \text{ Å} \times 110 \text{ Å} \times 30 \text{ Å}$) that brings E2–Ub bound to RBX1 into direct contact with CK1α (Fig. 4d). Mapping lenalidomide-induced CK1α ubiquitination sites identified Lys65 and Lys225 (numbering according to CK1α isoform 1) *in vivo*⁸ and Lys8, 62 and 179 *in vitro* (Extended Data Fig. 9b). These lysine residues are surface exposed and readily accessible to the E2 in our crystallographic model. These results illustrate, for the first time, how a cullin-RING ligase engages a large protein substrate, which leads to its ubiquitination by proximity to the E2. Numerous viral proteins make use of ubiquitination-by-proximity when hijacking the CRL1, CRL2, CRL4 or CRL5 systems to target antiviral host proteins for degradation²². Similarly, bifunctional chemical adaptors based on thalidomide-conjugated ligands (via the solvent-exposed phthalimide C4) were able to recruit diverse proteins to CRL4^{CRBN}, which lead to their degradation by proximity^{23,24}.

Interestingly, low molecular weight compounds have been shown to modulate protein–protein interactions²⁵ by either reinforcing pre-existing interactions²⁶ or inducing *de novo* complex formation²⁷. We did not detect binding of either IKZF1 or CK1α to CRL4^{CRBN} in the absence of IMiDs (Fig. 1b and Extended Data Figs 6a and 8c) and therefore, our results support an IMiD-induced *de novo* interaction between CK1α, IKZF1 and CRL4^{CRBN}. Recruitment of the CK1α and IKZF1 neo-substrates to CRL4^{CRBN} is driven by a surprisingly small 96.8 Å^2 compound–protein interface (phthalimide C5–C7) that induces high-affinity binding. It is currently unclear if IMiD-like metabolites or post-translational modifications exist²⁸ that assist recruitment of endogenous substrates through a related mechanism (Extended Data Fig. 9c).

Our work provides the first structural insight into how small molecules exploit an ubiquitin ligase for the destabilization of therapeutic targets, and offers a framework for designing future drugs that target proteins for proximity-based CRL4 ubiquitination and degradation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 September 2015; accepted 7 January 2016.

Published online 24 February 2016.

- Singhal, S. *et al.* Antitumor activity of thalidomide in refractory multiple myeloma. *N. Engl. J. Med.* **341**, 1565–1571 (1999).
- Bartlett, J. B., Dredge, K. & Dalglish, A. G. The evolution of thalidomide and its IMiD derivatives as anticancer agents. *Nature Rev. Cancer* **4**, 314–322 (2004).
- Ito, T. *et al.* Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350 (2010).
- Fischer, E. S. *et al.* Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49–53 (2014).
- Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
- Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014).
- Gandhi, A. K. *et al.* Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4^{CRBN}. *Br. J. Haematol.* **164**, 811–821 (2014).
- Krönke, J. *et al.* Lenalidomide induces ubiquitination and degradation of CK1 α in del(5q) MDS. *Nature* **523**, 183–188 (2015).
- List, A. *et al.* Efficacy of lenalidomide in myelodysplastic syndromes. *N. Engl. J. Med.* **352**, 549–557 (2005).
- Sekeres, M. A. & List, A. Lenalidomide (Revlimid, CC-5013) in myelodysplastic syndromes: is it any good? *Curr. Hematol. Rep.* **4**, 182–185 (2005).
- Schneider, R. K. *et al.* Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. *Cancer Cell* **26**, 509–520 (2014).
- Chamberlain, P. P. *et al.* Structure of the human Cereblon–DDB1–lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nature Struct. Mol. Biol.* **21**, 803–809 (2014).
- Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl Acad. Sci. USA* **103**, 17783–17788 (2006).
- Schjerven, H. *et al.* Selective regulation of lymphopoiesis and leukemogenesis by individual zinc fingers of Ikaros. *Nature Immunol.* **14**, 1073–1083 (2013).
- Zhou, S. *et al.* Transport of thalidomide by the human intestinal caco-2 monolayers. *Eur. J. Drug Metab. Pharmacokinet.* **30**, 49–61 (2005).
- Hofmeister, C. C. *et al.* Phase I trial of lenalidomide and CCI-779 in patients with relapsed multiple myeloma: evidence for lenalidomide–CCI-779 interaction via P-glycoprotein. *J. Clin. Oncol.* **29**, 3427–3434 (2011).
- Krenn, M., Gamcsik, M. P., Vogelsang, G. B., Colvin, O. M. & Leong, K. W. Improvements in solubility and stability of thalidomide upon complexation with hydroxypropyl-beta-cyclodextrin. *J. Pharm. Sci.* **81**, 685–689 (1992).
- Roche, S. *et al.* Development, validation and application of a sensitive LC-MS/MS method for the quantification of thalidomide in human serum, cells and cell culture medium. *J. Chromatogr. B* **902**, 16–26 (2012).
- Fischer, E. S. *et al.* The molecular basis of CRL4^{DDB2/CSA} ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024–1039 (2011).
- Angers, S. *et al.* Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery. *Nature* **443**, 590–593 (2006).
- Duda, D. M. *et al.* Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* **134**, 995–1006 (2008).
- Mahon, C., Krogan, N. J., Craik, C. S. & Pick, E. Cullin E3 ligases and their rewiring by viral factors. *Biomolecules* **4**, 897–930 (2014).
- Winter, G. E. *et al.* Phthalimide conjugation as a strategy for *in vivo* target protein degradation. *Science* **348**, 1376–1381 (2015).
- Lu, J. *et al.* Hijacking the E3 ubiquitin ligase cereblon to efficiently target BRD4. *Chem. Biol.* **22**, 755–763 (2015).
- Milroy, L. G., Grossmann, T. N., Hennig, S., Brunsvel, L. & Ottmann, C. Modulators of protein-protein interactions. *Chem. Rev.* **114**, 4695–4748 (2014).
- Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640–645 (2007).
- Griffith, J. P. *et al.* X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell* **82**, 507–522 (1995).
- Hartmann, M. D. *et al.* Thalidomide mimics uridine binding to an aromatic cage in cereblon. *J. Struct. Biol.* **188**, 225–232 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Novartis Research Foundation, the Swiss National Science Foundation (31003A_144020) and the European Research Council (ERC-2014-ADG 666068 CsnCRL). G.P. was supported by Long-Term Fellowships of the European Molecular Biology Organization (EMBO; ALTF-1350-2013) and the Human Frontier Science Program (HFSP; LT000210/2014). We thank G. M. Lingaraju for the DDB1 ^{Δ BPB} construct, M. Faty for assistance with protein purifications, J. Rabl and S. Kassube for help during diffraction data collection, R. Bunker for support with data processing and model building, W. A. Rahman and A. Potenza for assistance with fluorescence polarization, W. C. Forrester for help and comments, and J. Seebacher, R. Sack, D. Hess and M. Stadler for help with mass spectrometry experiments. We acknowledge the Paul Scherrer Institute for provision of synchrotron radiation beam time at beamline PXII and PXIII of the SLS and would like to thank M. Wang for assistance.

Author Contributions G.P., E.S.F. and N.H.T. designed the experiments and interpreted the results. G.P. and E.S.F. carried out the experiments. G.P., E.S.F. and N.H.T. wrote the manuscript.

Author Information Structural coordinates for hsDDB1 ^{Δ BPB}–hsCRBN–lenalidomide–hsCK1 α have been deposited in the Protein Data Bank under accession number 5FQD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.H.T. (nicolas.thoma@fmi.ch).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Constructs and protein purification. Wild-type and mutant versions of human DDB1 (Q16531), human CRBN (Q96SW2), human CK1 α (P48729) and human IKZF1 (Q13422) were subcloned into pAC-derived vectors²⁹ and recombinant proteins expressed as N-terminal His₆, StrepII or StrepII-Avi fusions in *Trichoplusia ni* High-Five insect cells using the baculovirus expression system (Invitrogen). For purification of His₆-DDB1-His₆-CRBN, His₆-DDB1 Δ ^{BPB}-StrepII-CRBN Δ ^{1–40} and His₆-DDB1 Δ ^{BPB}-His₆-CRBN, cells were resuspended in buffer containing 50 mM Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) pH 8.0, 200 mM NaCl, 0.25 mM Tris(2-carboxyethyl)phosphine (TCEP), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 \times protease inhibitor cocktail (Sigma-Aldrich) and lysed by sonication. Cells expressing StrepII-Avi-CK1 α or truncated versions of StrepII-Avi-IKZF1 (Δ 256–519, Δ 197–238/ Δ 256–519 and Δ 1–82/ Δ 197–238/ Δ 256–519; full-length IKZF1 forms aggregates during purification) were lysed in the presence of 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.25 mM TCEP, 1 mM PMSF and 1 \times protease inhibitor cocktail (Sigma-Aldrich). Following ultracentrifugation, the soluble fraction was passed over Strep-Tactin Sepharose (IBA) or His-Select nickel affinity resin (Sigma-Aldrich) and following elution, affinity tags were removed from CK1 α and IKZF1 by overnight TEV protease treatment as indicated. The affinity-purified protein was further purified via ion exchange chromatography (Poros 50HQ and 50HS) and subjected to size-exclusion chromatography in 50 mM HEPES pH 7.4, 200 mM NaCl and 0.25 mM TCEP. The protein-containing fractions were concentrated using ultrafiltration (Millipore) and flash frozen (DDB1–CRBN constructs at 40–120 μ M, TEV cleaved CK1 α at \sim 280 μ M, StrepII-Avi-CK1 α at \sim 50 μ M, TEV cleaved IKZF1 at \sim 350 μ M and StrepII-Avi-IKZF1 at \sim 150 μ M). Proteins were stored at -80°C .

Crystallization and data collection. Attempts to crystallize CK1 α , lenalidomide and CRBN with the full-length DDB1 adaptor protein were unsuccessful. Because the WD40 β -propeller B (BPB) of DDB1 is not involved in CRBN binding, we generated a human DDB1 construct in which a GNGNSG-linker replaced the BPB domain (residues 396–705; DDB1 Δ ^{BPB}). For crystallization of the DDB1 Δ ^{BPB}-CRBN–lenalidomide–CK1 α complex, His₆-DDB1 Δ ^{BPB}-StrepII-CRBN Δ ^{1–40} at 70 μ M was mixed with lenalidomide at 80 μ M before the addition of TEV cleaved full-length CK1 α at 80 μ M. The mixture was incubated on ice for 1 h and subsequently centrifuged at 20,000g for 30 min at 4°C . Crystallization plates were set up and stored at room temperature. Crystals appeared within 3 days after mixing the protein solution 1:1 with the reservoir containing 70 mM Tris pH 7.0, 140 mM MgCl₂ and 7% (w/v) PEG 8000 and continued growing until day 13 in MRC 2 Well Crystallization format vapour diffusion plates (Swissci). Crystals were cryo-protected in reservoir solution supplemented with 20% ethylene glycol and flash-cooled in liquid nitrogen. Diffraction data were collected at the Swiss Light Source (beamline PXII) with a Pilatus 6M detector at the wavelength of the Zn-edge (1.28162 Å) and a temperature of 100 K. Data were indexed and integrated using XDS³⁰ and scaled using AIMLESS supported by other programs of the CCP4 suite³¹. The optimal high-resolution cut-off (2.45 Å) was determined based on the half-set correlation criterion³² ($CC_{1/2} = 0.45$ for the highest resolution shell). Data processing statistics, refinement statistics and model quality parameters are provided in Extended Data Table 1.

Structure determination and model building. The DDB1 Δ ^{BPB}-CRBN–lenalidomide–CK1 α quaternary complex crystallized in space group *P1* with two complexes in the unit cell. PHASER³³ was used to determine the structure by molecular replacement using a crystallographic model of DDB1 omitting the BPB domain, a CK1 α homology model generated with Modeller³⁴ based on a CK1 δ crystal structure (PDB entry 4TWC), and human CRBN (PDB entry 4TZ4) as search models. The initial model was iteratively improved with COOT and refined using PHENIX.REFINE³⁵ and autoBUSTER³⁶, with ligand restraints generated by Grade server (Global Phasing). Figures were generated with PyMOL (The PyMOL Molecular Graphics System, Version 1.7.2.2 Schrödinger, LLC) and model quality was assessed with MOLPROBITY³⁷. Interaction surfaces were determined with PISA³⁸. The IKZF1 homology model was calculated using Modeller based on a multiple sequence alignment with the experimental zinc-finger structures of Aart, YY1 and Kasio (PDB entries 2I13, 1UBD, 2LT7).

Labelling of NEDD8 with Alexa-488–maleimide. Purified human NEDD8(M1C) (NEDD8) was incubated with DTT (8 mM) at 4°C for 1 h. DTT was removed using a S200 16/60 gel filtration column in a buffer containing 50 mM Tris pH 7.3 and 150 mM NaCl. Alexa-488–C5–maleimide (Invitrogen) was dissolved in 100% DMSO and mixed with NEDD8 to achieve fourfold molar excess of Alexa-488–C5–maleimide. NEDD8 labelling was carried out at room temperature for 3 h in a vacuum desiccator and stored overnight at 4°C . Labelled NEDD8 was purified on a S200 16/60 gel filtration column in 50 mM Tris pH 7.5, 150 mM NaCl, 0.25 mM

TCEP and 10% (v/v) glycerol, concentrated by ultrafiltration (Millipore), flash frozen (-40 – -80°C) in liquid nitrogen and stored at -80°C .

CRL4^{CRBN} reconstitution and *in vitro* CUL4A neddylation with Alexa-488–NEDD8. *In vitro* CRL4^{CRBN} reconstitution and CUL4A neddylation was performed as described^{4,19,21}. His₆-CUL4A–His₆-RBX1 at 3.5 μ M was incubated with His₆-DDB1–His₆-CRBN at 1.5 μ M (wild-type or mutant forms) in a reaction mixture containing 3.8 μ M Alexa-488–NEDD8, 50 nM NAE1/UBA3 (E1), 150 nM UBC12 (E2), 1 mM ATP, 50 mM Tris pH 7.5, 100 mM NaCl, 2.5 mM MgCl₂, 0.5 mM DTT and 5% (v/v) glycerol for 2 h at room temperature. Gel filtration purified neddylated CRL4^{CRBN} ($\text{N}_8\text{CRL4}^{\text{CRBN}}$) was concentrated to 2–4 μ M, flash frozen and stored at -80°C . Alexa-488–NEDD8 labelling efficiency of different CRL4^{CRBN} mutant complexes was determined by measuring total fluorescent intensity at equal concentration (excitation at 490 nm, emission at 540 nm) using a Safire2 microplate reader from Tecan (Extended Data Fig. 4f).

Biotinylation of CK1 α and IKZF1. Purified StrepII-Avi-tagged CK1 α or IKZF1 were biotinylated *in vitro* at a concentration of 25–50 μ M by incubation with final concentrations of 2.5 μ M BirA enzyme and 0.2 mM D-Biotin in 50 mM HEPES pH 7.4, 200 mM NaCl, 10 mM MgCl₂, 0.25 mM TCEP and 20 mM ATP. The reaction was incubated for 1 h at room temperature and stored at 4°C for 14–16 h. Biotinylated proteins were purified by gel filtration chromatography and stored at -80°C (StrepII-Avi-CK1 α at \sim 25 μ M, StrepII-Avi-IKZF1 at \sim 40 μ M).

Time-resolved fluorescence resonance energy transfer (TR-FRET). Increasing concentrations of Alexa-488–NEDD8-labelled CRL4^{CRBN} ($\text{N}_8\text{CRL4}^{\text{CRBN}}$) were added to pre-mixed biotinylated IKZF1 at 80 nM or CK1 α at 100 nM, terbium-coupled streptavidin at 4 nM (Invitrogen) and IMiDs or 2'-deoxyuridine at 5 μ M (final concentrations) in 384-well microplates (Greiner, 784076) in a buffer containing 50 mM Tris pH 7.5, 100 mM NaCl, 0.1% pluronic acid and 1% DMSO (see also figure legends). Before TR-FRET measurements were conducted, the reactions were incubated for 15 min at room temperature. After excitation of terbium (Tb) fluorescence at 337 nm, emission at 490 nm (Tb) and 520 nm (Alexa 488) were recorded with a 70 μ s delay to reduce background fluorescence and the reaction was followed over 1 h by recording 60 technical replicates of each data point using a PHERAstar FS microplate reader (BMG Labtech). The TR-FRET signal of each data point was extracted by calculating the 520/490 nm ratio. Data were analysed with GraphPad Prism 6 assuming equimolar binding of the probe to the receptor ($\text{N}_8\text{CRL4}^{\text{CRBN}}$) using the following equations:

The concentration of the receptor in the bound state, $[C_{\text{Rbound}}]$, can be calculated for that setting by the law of mass action:

$$[C_{\text{Rbound}}] = (K_D + [C_{\text{Ptot}}] + [C_{\text{Rtot}}] - ((K_D + [C_{\text{Ptot}}] + [C_{\text{Rtot}}])^2 - 4 \times [C_{\text{Ptot}}] \times [C_{\text{Rtot}}])^{0.5})/2$$

K_D is the equilibrium constant for the dissociation, and $[C_{\text{Ptot}}]$ and $[C_{\text{Rtot}}]$ are the total concentrations of the probe and the receptor, respectively.

The K_D value can be calculated from the change in the fluorescence intensity, FI , observed by a titration of the receptor at constant probe concentrations according to:

$$FI_{\text{obs}} = FI_{\text{free}} + (FI_{\text{bound}} - FI_{\text{free}}) \times (K_D + [C_{\text{Ptot}}] + [C_{\text{Rtot}}] - ((K_D + [C_{\text{Ptot}}] + [C_{\text{Rtot}}])^2 - 4 \times [C_{\text{Ptot}}] \times [C_{\text{Rtot}}])^{0.5})/(2 \times [C_{\text{Rtot}}])$$

FI_{obs} is the observed fluorescence intensity, and FI_{free} and FI_{bound} the fluorescence intensities of the probe in its free and its bound states, respectively. The assay window is described by the overall change in the fluorescence intensity, $(FI_{\text{bound}} - FI_{\text{free}})$.

Counter titrations with unlabelled proteins were carried out by mixing $\text{N}_8\text{CRL4}^{\text{CRBN}}$ at 0.5–1 μ M with 200 nM biotinylated CK1 α or 160 nM biotinylated IKZF1 in the presence of 8 nM terbium-coupled streptavidin and IMiDs at 10–20 μ M. After 15 min incubation on ice, increasing amounts of unlabelled DDB1 Δ ^{BPB}-CRBN (0.04–40 μ M) or IKZF1 (0.04–20 μ M; wild-type or mutant forms or pre-incubated with consensus or control DNA as stated in the EMSA section) were added to the pre-assembled $\text{N}_8\text{CRL4}^{\text{CRBN}}$ -CK1 α /IKZF1 complexes in a 1:1 volume ratio and incubated for 5 min at room temperature. IMiD titrations were carried out by premixing $\text{N}_8\text{CRL4}^{\text{CRBN}}$, CK1 α /IKZF1 and 8 nM terbium-coupled streptavidin before addition of increasing concentrations of each IMiD (0.005–5 μ M) in a 1:1 volume ratio (see figure legends for final concentrations). The 520/490 nm ratios were plotted to calculate the half maximal effective concentrations (EC_{50}) or half maximal inhibitory concentrations (IC_{50}) assuming a single binding site using GraphPad Prism 6. IC_{50} values were converted to the respective K_i as described³⁹. Three biological replicates were carried out per experiment. The standard deviation was derived from the sum of the mean absolute error of 10 technical replicates (per data point and replicate) and the standard deviation of the biological replicates.

Fluorescence polarization to monitor binding of CRBN mutants to IMiDs. Cy5-conjugated thalidomide⁴ (10 nM) was mixed with increasing concentrations of either wild-type or mutant forms of purified DDB1–CRBN (0.004–2 μ M) in a 384-well microplate (Greiner, 784076) and incubated for 30 min at room temperature. CRBN–ImiD interactions were measured in 50 mM Tris pH 7.5, 100 mM NaCl, 0.1% pluronic acid and 1% DMSO by change in fluorescence polarization using a PHERAstar FS microplate reader (BMG Labtech). The Cy5-thalidomide bound fraction was calculated as described⁴⁰. Data were plotted and analysed using GraphPad Prism 6 assuming a single ImiD binding site on CRBN.

Label-free quantification mass spectrometry. Hct116 cells were purchased from The European Collection of Cell Cultures (ECACC, Sigma-Aldrich), immediately used for experiments and regularly tested for mycoplasma contamination. Cells were cultured in L-lysine and L-arginine free DMEM supplemented with unlabelled L-lysine, L-arginine, 10% FBS and 2 mM L-glutamine. Cells were grown to approximately 50% confluency and the medium was exchanged for DMEM supplemented with ¹³C,¹⁵N-labelled L-lysine (Lys8) and ¹³C,¹⁵N-labelled L-arginine (Arg10) containing lenalidomide at 30 μ M or equivalent amounts of DMSO as control. Cells were incubated for 16 h and harvested for mass spectrometry analysis in 0.5 M Tris-HCl pH 8.6, 6 M guanidine hydrochloride, reduced in 16 mM TCEP for 30 min, and alkylated in 35 mM iodoacetamide for 30 min in the dark. The proteins were digested at 37 °C with lysyl endopeptidase (Wako) after dilution to ~2 M guanidine hydrochloride (with 50 mM Tris-HCl pH 7.3, 5 mM CaCl₂ buffer) for 6 h, and after dilution to <1 M guanidine hydrochloride with trypsin (Promega) at 37 °C overnight. The resulting peptides were desalted using C₁₈ solid state extraction cartridges and offline fractionated into 36 fractions by basic reverse phase chromatography. The 36 fractions were recombined to a final of 12 samples.

Generated peptides were separated on an EASY n-LC 1000 liquid chromatography system equipped with a C₁₈ EASY-Spray column coupled to an Orbitrap Fusion mass spectrometer (all from Thermo Scientific).

Maxquant⁴¹ was used for .RAW file processing and controlling peptide and protein level false-discovery rates, assembling proteins from peptides, and protein quantification from peptides. Peptides were searched against a human Uniprot database with both forward and reverse sequences. For analysis, we did not utilize the SILAC component of the experiment, but instead used the Maxquant LFQ algorithm to quantify the relative abundance of casein kinase isoforms across three independent replicates.

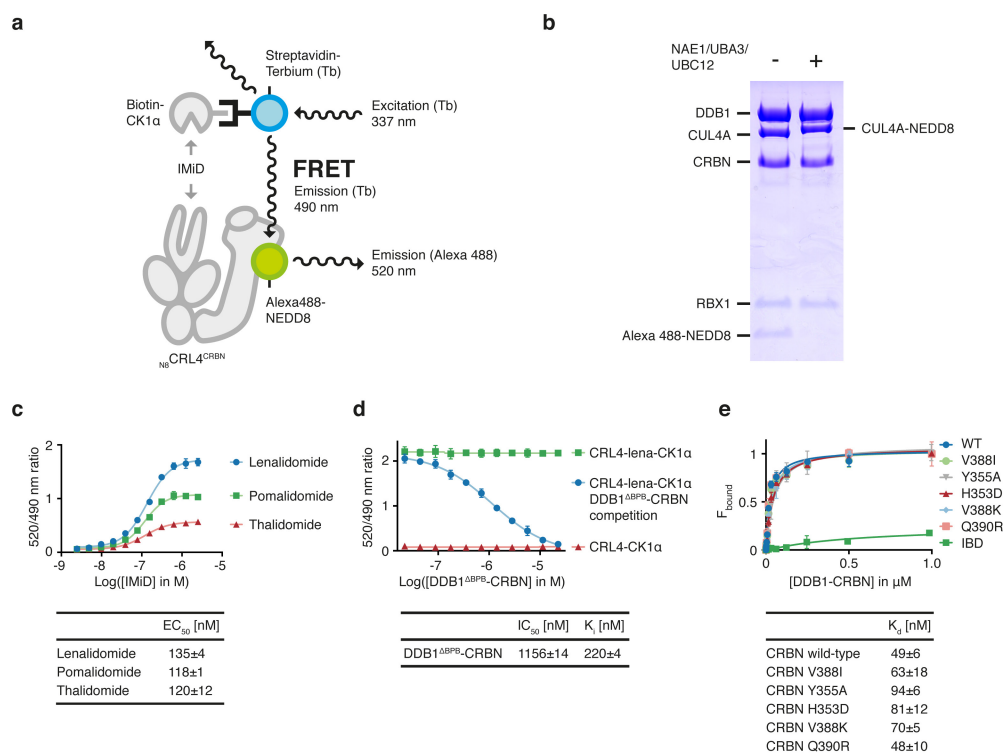
In vitro ubiquitination and identification of target lysines. *In vitro* ubiquitination was performed by mixing wild-type or mutant N⁸CRL4^{CRBN} at 70 nM with a reaction mixture containing IMiDs at 350 nM or 10 μ M, CK1 α at 500 nM, E1 (UBA1, BostonBiochem) at 40 nM, E2 (UBCH5a, BostonBiochem) at 1 μ M, wild-type (20 μ M) or lysine-free (10 μ M) ubiquitin as indicated. Reactions were carried out in 50 mM Tris pH 7.5, 30 mM NaCl, 5 mM MgCl₂, 0.2 mM CaCl₂, 1 mM ATP, 0.1% Triton X-100 and 0.1 mg ml⁻¹ BSA, incubated for 15–30 min at 30 °C and analysed by western blot using anti-CK1 α (abcam, ab108296, 1:20,000) and anti-rabbit IRDye 800CW antibodies (LI-COR, 926-32211, 1:10,000). Blots were scanned on a LI-COR Odyssey infrared imaging system.

To identify target lysines following *in vitro* ubiquitination of CK1 α in the presence or absence of lenalidomide, samples were precipitated with 20% (v/v)

trichloroacetic acid followed by several acetone washes. The precipitated protein was dissolved in 10 μ l of 500 mM Tris pH 8.6, 6 M guanidine hydrochloride and 8 mM TCEP and incubated with 18 mM iodoacetamide for 30 min at room temperature in the dark. After addition of 50 μ l 50 mM Tris pH 7.4, 5 mM CaCl₂, samples were incubated with 200 ng trypsin at 37 °C for 12–14 h. Mass spec was carried out on a FUSION Orbitrap, the data was searched with MASCOT, site probabilities were calculated with ASCOR and peak integration for relative quantification was performed using ProgenesisLC.

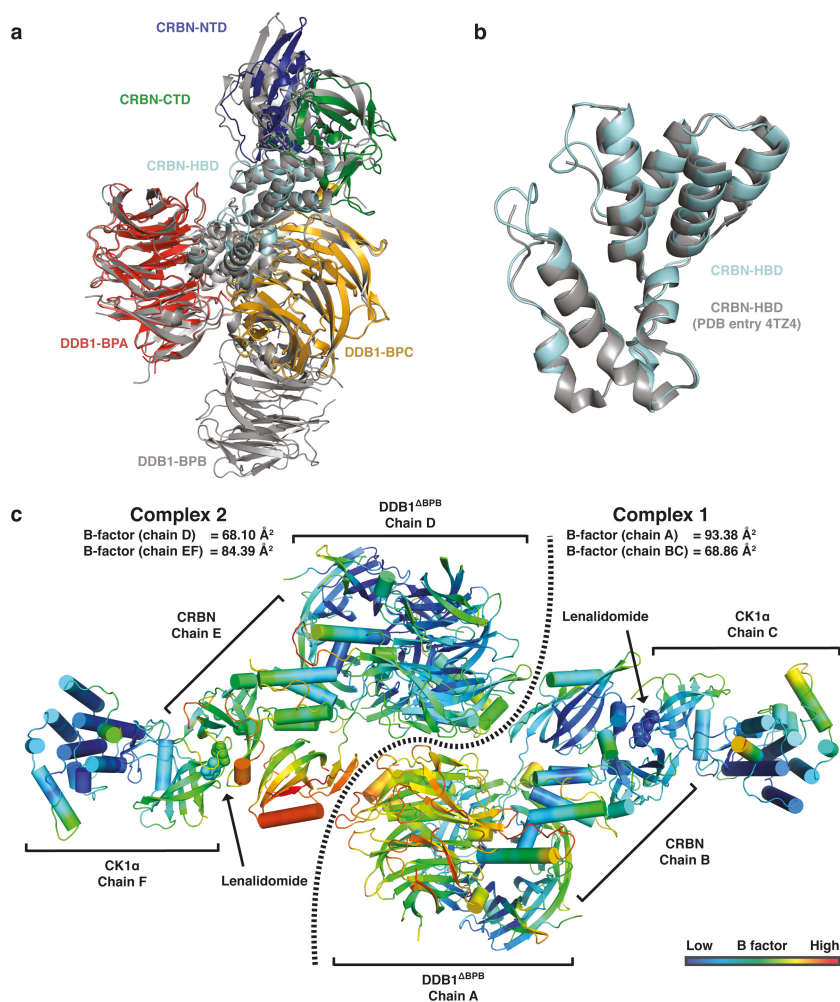
Electrophoretic mobility shift assays (EMSAs). Equimolar ratios of complementary DNA single strands at 200 mM were mixed in 10 mM Tris pH 8.0, 50 mM NaCl and 1 mM MgCl₂, incubated at 95 °C for 2 min and annealed by slowly lowering temperature using a thermal cycler. IKZF1 was mixed with duplex DNA (consensus DNA: 5'-TCAGAAAAAGGGAATTCCGTCAC-3'; control DNA: 5'-TCAGACACTTTTGGTACTGTAC-3') in 50 mM HEPES pH 7.4, 200 mM NaCl, 0.25 mM TCEP and 10% (v/v) glycerol and incubated on ice for 15 min before the addition of DDB1–CRBN and pomalidomide, as indicated. Binding reactions were incubated for 30 min at room temperature, applied to a 4–16% NativePAGE Novex Bis-Tris gel (Invitrogen) and separated in 1 \times NativePAGE buffer at 150 V for 75 min at room temperature. Gels were stained with 1 μ g ml⁻¹ ethidium bromide in 1 \times NativePAGE buffer followed by Coomassie staining.

29. Abdulrahman, W. *et al.* A set of baculovirus transfer vectors for screening of affinity tags and parallel expression strategies. *Anal. Biochem.* **385**, 383–385 (2009).
30. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
31. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
32. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
33. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
34. Šali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318–326 (1995).
35. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
36. Bricogne, G. B. E. *et al.* BUSTER version 2.11.5. (Global Phasing Ltd., 2011).
37. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
38. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
39. Cer, R. Z., Mudunuri, U., Stephens, R. & Lebeda, F. J. *IC₅₀-to-K_i*: a web-based tool for converting *IC₅₀* to *K_i* values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* **37**, W441–W445 (2009).
40. Marks, B. D. *et al.* Multiparameter analysis of a screen for progesterone receptor ligands: comparing fluorescence lifetime and fluorescence polarization measurements. *Assay Drug Dev. Technol.* **3**, 613–622 (2005).
41. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372 (2008).



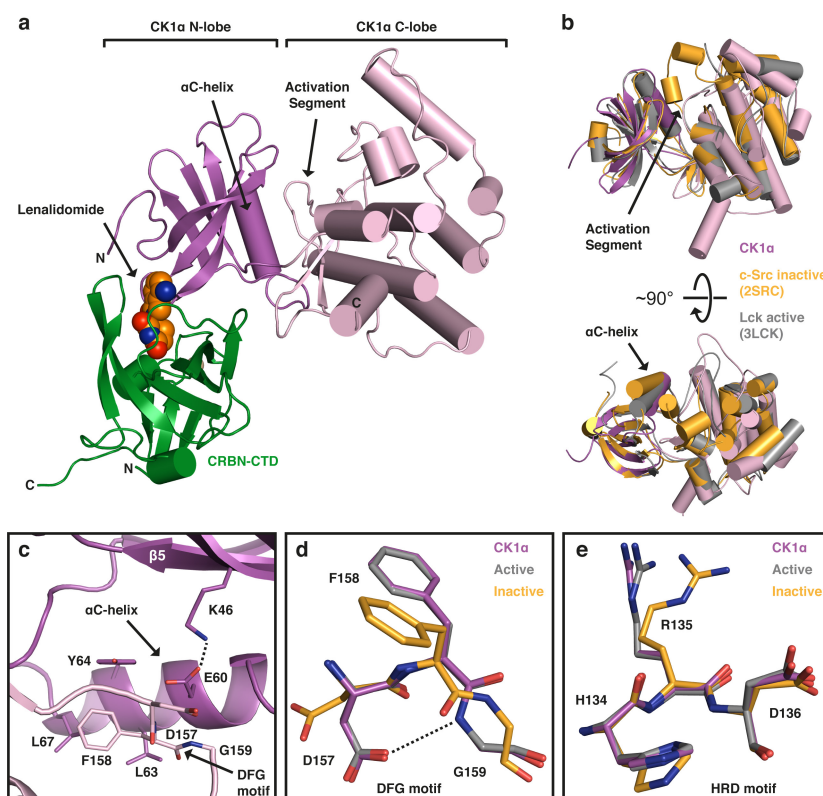
Extended Data Figure 1 | Characterization of the CRL4^{CRBN}-CK1 α interaction. **a**, TR-FRET setup. **b**, Neddylation of CRL4^{CRBN} with Alexa 488-labelled NEDD8. **c**, Titration of lenalidomide, pomalidomide and thalidomide (0.0025–2.5 μ M) to N₈CRL4^{CRBN} at 300 nM and CK1 α at 100 nM. EC_{50} values show that similar compound concentrations are required to bind CK1 α to N₈CRL4^{CRBN}. **d**, Unlabelled DDB1 ΔBPB -CRBN competes with N₈CRL4^{CRBN} for CK1 α binding. Titration of unlabelled wild-type DDB1 ΔBPB -CRBN (0.02–20 μ M) to N₈CRL4^{CRBN}-lenalidomide-CK1 α complexes pre-assembled by mixing N₈CRL4^{CRBN} at 450 nM, CK1 α

at 100 nM and lenalidomide at 10 μ M (blue curve). Addition of buffer to N₈CRL4^{CRBN}-lenalidomide-CK1 α (green curve; maximum signal) and N₈CRL4^{CRBN}, CK1 α without lenalidomide (red curve; minimum signal). IC_{50} was converted to the respective K_i as described in methods. The resulting K_i is similar to the K_D as determined by non-competitive TR-FRET (Fig. 1a). **e**, Binding of Cy5-labelled thalidomide to CRBN mutants (DDB1-CRBN) determined by fluorescence polarization. Data in this figure are biological replicates presented as means \pm s.d. ($n = 3$).



Extended Data Figure 2 | The asymmetric unit contains two quaternary complexes. a, DDB1^{ΔBPB}-CRBN superimposes on DDB1-CRBN (PDB entry 4TZ4; grey) with an overall r.m.s.d. of 0.96 Å. **b,** CRBN helical bundle domains (HBD) extracted from **a**. The CRBN-HBDs superimpose with an overall r.m.s.d. of 0.44 Å. **c,** The DDB1^{ΔBPB}-CRBN-lenalidomide-CK1α

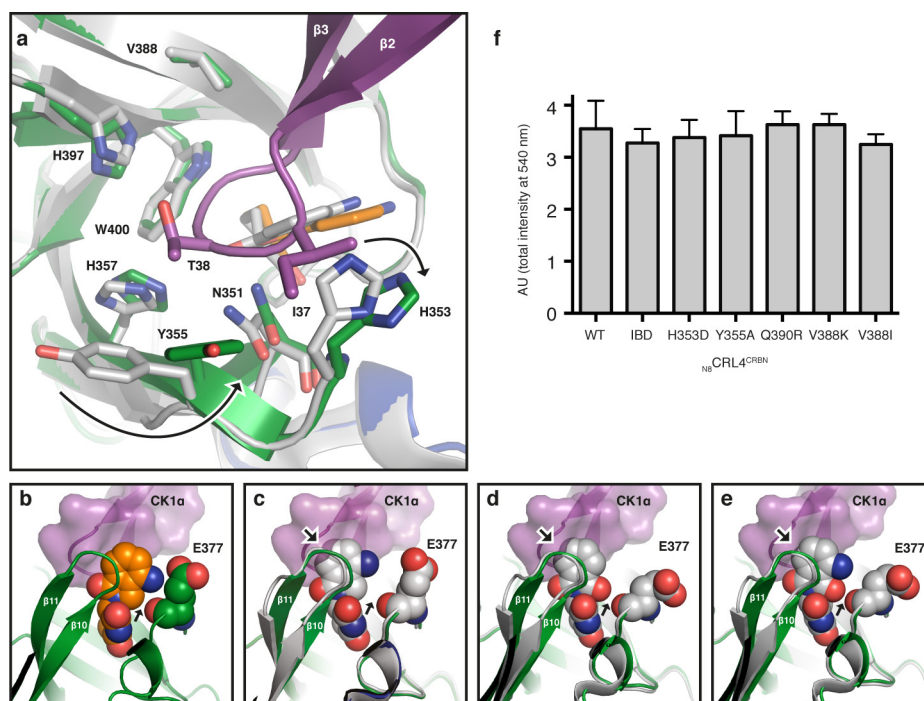
crystal contained two quaternary complexes in the unit cell. The two complexes are shown as cartoon representation coloured according to B factor. Lenalidomide is shown as spheres. Average B-factors for different parts of the molecule are provided.



Extended Data Figure 3 | CK1α adopts an active kinase conformation.

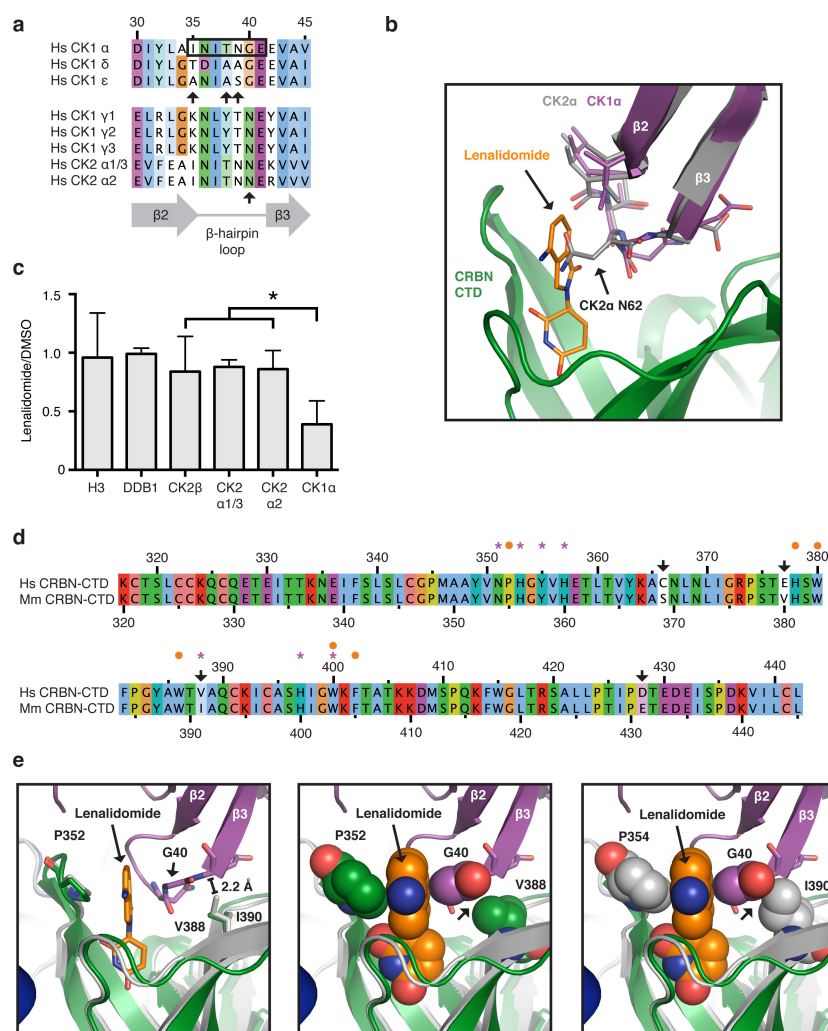
a, C-terminal domain of CRBN, lenalidomide and CK1α. Elements required for kinase activation are indicated. **b**, Superposition of CK1α with inactive c-Src kinase (PDB entry 2SRC; orange) and an active conformation of human lymphocyte kinase Lck (PDB entry 3LCK; grey) shown in two different orientations. Conformational differences in the activation segment and the αC-helix are indicated. **c**, Close-up of the

αC-helix of CK1α. The DFG motif makes hydrophobic contacts with the αC-helix. The Lys–Glu hydrogen bond (dashed line) is characteristic of an active kinase conformation. **d**, Superposition of the DFG motifs of CK1α, c-Src and Lck. The CK1α DFG motif adopts an active conformation. **e**, Superposition of the HRD motifs of CK1α, c-Src and Lck. The CK1α HRD motif adopts an active conformation.



Extended Data Figure 4 | CK1 α binding induces minor changes in CRBN. **a**, Superposition of DDB1 Δ^{BPB} -CRBN-lenalidomide-CK1 α (coloured as in Fig. 1c) with human DDB1-CRBN (PDB 4TZ4) in grey. Only the CRBN residues His353 and Tyr355 undergo minor rearrangements upon CK1 α binding. **b**, Lenalidomide orientation (shown as spheres) in the DDB1 Δ^{BPB} -CRBN-lenalidomide-CK1 α complex. C3 of the phthalimide group is close to the main chain carbonyl group of Glu377 (shown as spheres). **c**, Orientation of lenalidomide in the structure of human DDB1-CRBN (PDB entry 4TZ4; grey) superimposed with DDB1 Δ^{BPB} -CRBN-lenalidomide-CK1 α . Clashes of phthalimide C6

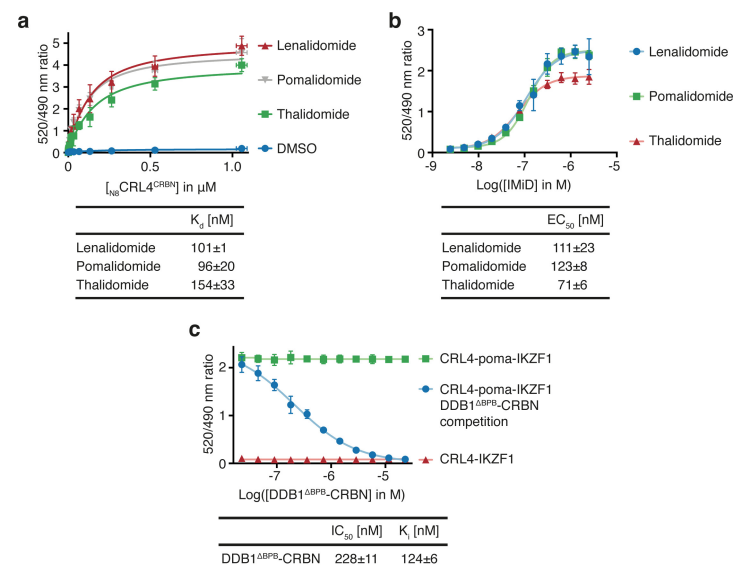
and C7 with CK1 α are indicated by an arrow. **d**, **e**, Thalidomide (**d**) and pomalidomide (**e**) in the structure of *Homo sapiens* DDB1 and *Gallus gallus* CRBN (PDB entry 4CI1 and 4CI3) in grey. The β -strands 10 and 11 are in an identical conformation as compared to lenalidomide-bound human DDB1-CRBN (PDB entry 4TZ4) and are not differentially regulated by binding to different IMiDs⁸. **f**, Total intensity measurements of different Alexa-488-labelled N_8CRL4^{CRBN} mutant complexes confirms comparable labelling efficiency (AU, arbitrary units). Data are biological replicates presented as means \pm s.d. ($n = 3$).



Extended Data Figure 5 | Conservation of the CRBN-CK1 α interface.

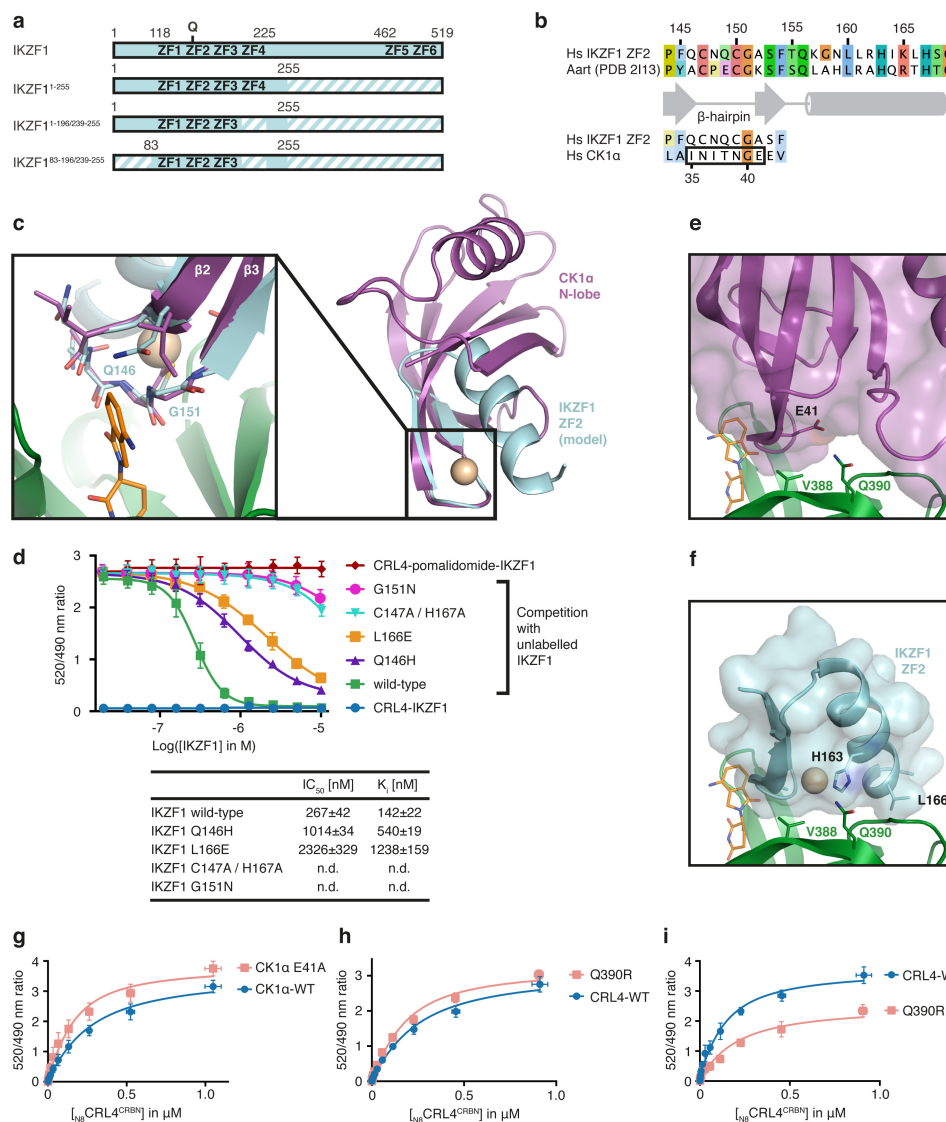
a, Sequence of the interacting CK1 α β -hairpin loop aligned to different CK1 and CK2 isoforms. **b**, Superposition of the CK1 α and CK2 α (PDB entry 3U87) β -hairpin loops. CK2 α residue Asn62 aligns with CK1 α residue Gly40 and clashes with lenalidomide. **c**, Quantification of CK1 α and CK2 α cellular abundance using label-free quantification mass spectrometry. Two-tailed paired *t*-test analysis revealed significant differences between CK2 β , CK2 α 1/3, CK2 α 2 and CK1 α with $P < 0.05$. Data are biological replicates presented as means \pm s.d. ($n = 3$). **d**, Sequence alignment of human and mouse CRBN C-terminal domains.

Asterisks indicate residues that contact CK1 α . Orange dots mark residues required for IMiD binding. Arrows highlight the four diverging residues. Human CRBN residues C366 and D428 are not part of the CRBN-CK1 α interface. **e**, Superposition of CRBN-lenalidomide-CK1 α (coloured as in Fig. 1c) with chicken CRBN (PDB entry 4CI2) in grey. CRBN Pro352 and CK1 α Gly40 make close contacts with the phthalimide ring of lenalidomide. CRBN Val388 is close to CK1 α Gly40 (arrow; middle panel). Substitution of Val388 for isoleucine leads to clashes with CK1 α , as illustrated by chicken CRBN Ile390 (arrow; right panel).



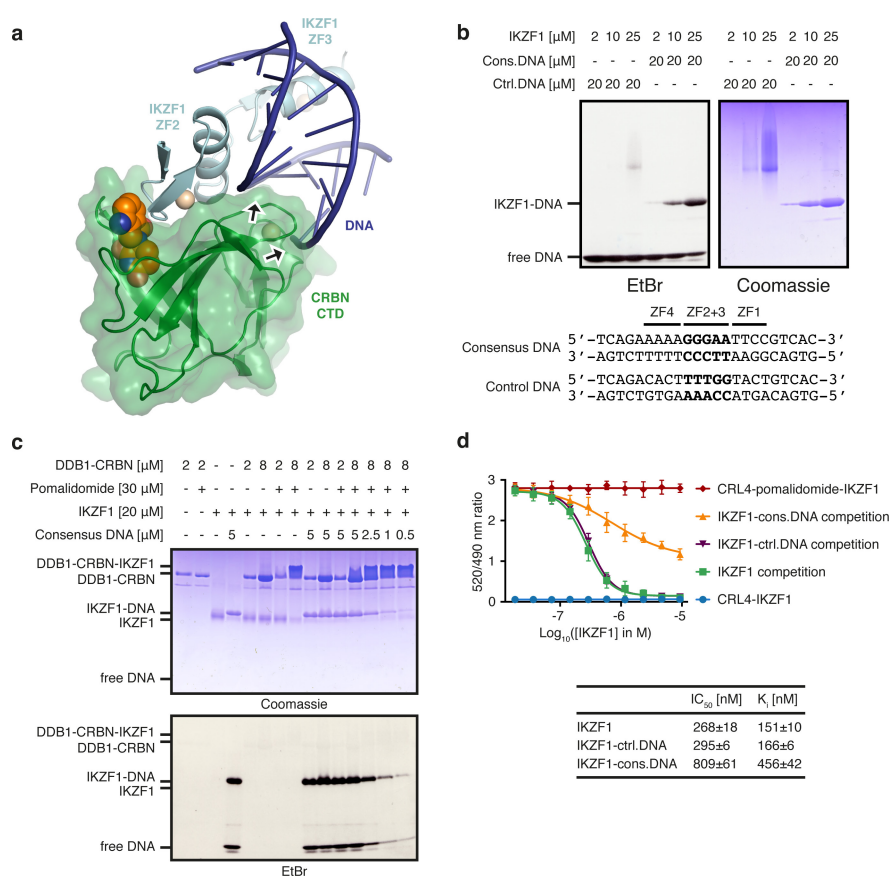
Extended Data Figure 6 | Characterization of the CRL4^{CRBN}–IKZF1 interaction. **a**, N₈CRL4^{CRBN}–IKZF1 complex formation measured by TR-FRET in the absence (DMSO) or presence of lenalidomide, pomalidomide or thalidomide at saturating concentrations (5 μ M). **b**, Titration of lenalidomide, pomalidomide and thalidomide (0.0025–2.5 μ M) to N₈CRL4^{CRBN} at 200 nM and IKZF1 at 80 nM. EC₅₀ values show that equal compound concentrations are required to bind IKZF1 to N₈CRL4^{CRBN}. **c**, Titration of unlabelled wild-type DDB1^{ΔBPP}–CRBN (0.02–20 μ M)

to N₈CRL4^{CRBN}–lenalidomide–IKZF1 complexes pre-assembled by mixing N₈CRL4^{CRBN} at 250 nM, IKZF1 at 70 nM and pomalidomide at 10 μ M. Addition of buffer to N₈CRL4^{CRBN}–pomalidomide–IKZF1 (green curve; maximum TR-FRET signal) and N₈CRL4^{CRBN}, IKZF1 without pomalidomide (red curve; minimum TR-FRET signal). IC₅₀ values were converted to the respective K_i as described in methods. Data in this figure are biological replicates presented as means \pm s.d. ($n = 3$).



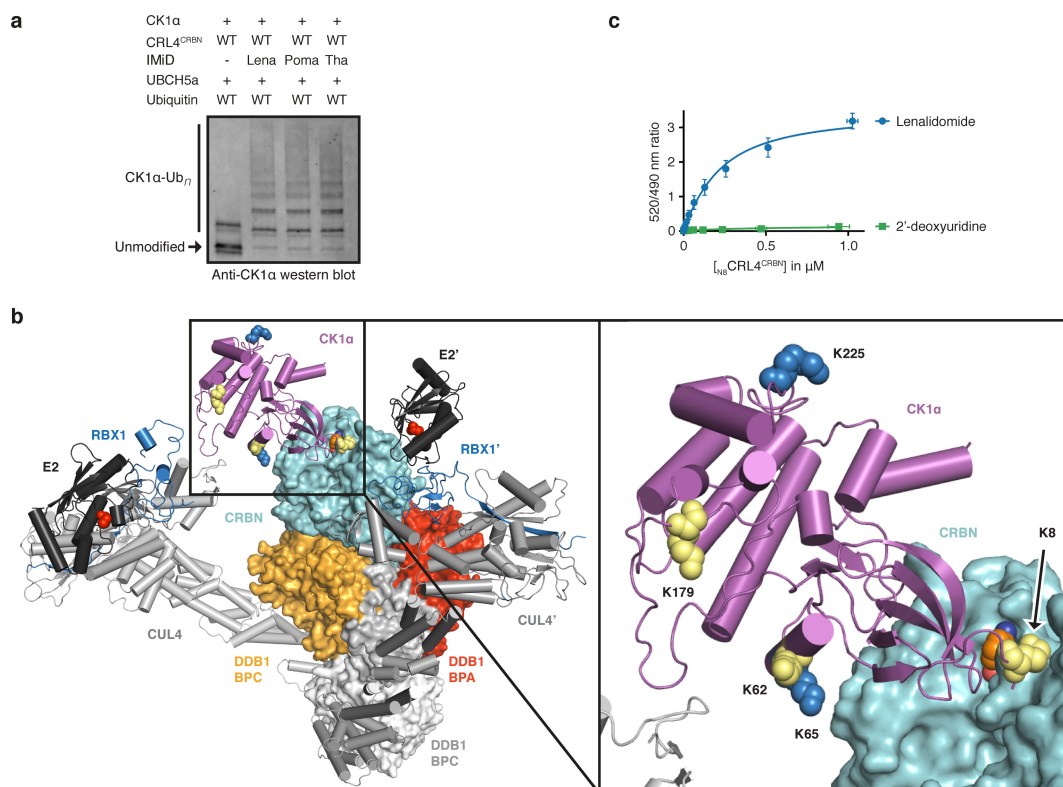
Extended Data Figure 7 | IKZF1 utilizes a β -hairpin loop to bind CRBN. **a**, Overview of IKZF1 constructs. ZF, zinc finger; Q, Gln146. Hatched lines indicate omitted sequences. **b**, Sequence alignment between human IKZF1-ZF2, the β -hairpin loop of human CK1 α and Aart (PDB 2I13; residues 133–158). PDB entries 2I13, 2LT7 and 1UBD were used for IKZF1 modelling. **c**, Superposition of modelled IKZF1-ZF2 with the CK1 α β -hairpin loop. **d**, Unlabelled wild-type or mutant IKZF1^{83–196/239–255} (0.02–10 μ M) added to N₈CRL4^{CRBN}-pomalidomide-IKZF1 complexes pre-assembled by mixing N₈CRL4^{CRBN} at 250 nM, IKZF1^{1–196/239–255} at 80 nM and pomalidomide at 5 μ M. N₈CRL4^{CRBN}-

pomalidomide-IKZF1 complex (red curve; maximum TR-FRET signal). N₈CRL4^{CRBN}, IKZF1 without pomalidomide (blue curve; minimum TR-FRET signal). **e**, CBRN-CK1 α interface. **f**, Model of the CBRN-IKZF1 interface. CBRN Gln390 is proximal to IKZF1 His163 in the α -helix of ZF2. **g**, Mutation of CK1 α Glu41 to alanine, the corresponding residue in IKZF1, strengthens binding ($K_D = 109 \pm 30$ nM). **h**, Mutation of CBRN Gln390 to arginine has no substantial influence on CK1 α binding, but interferes with IKZF1 recruitment (**i**) as predicted by the model. TR-FRET data in this figure are biological replicates presented as means \pm s.d. ($n = 3$).



Extended Data Figure 8 | Presence of IKZF1 consensus DNA interferes with binding to CRBN. **a**, Model of IKZF1 bound to DNA and CRBN. DNA clashes with the zinc-binding region of CRBN. **b**, Electrophoretic mobility shift assay. IKZF1^{83-196/239-255} specifically binds its consensus DNA (cons. DNA). **c**, Electrophoretic mobility shift assay. Presence of DNA affects the interaction between DDB1-CRBN, pomalidomide and IKZF1. DNA does not comigrate with the DDB1-CRBN-IKZF1 complex. DDB1-CRBN and IKZF1 do not interact in the absence of pomalidomide. **d**, IKZF1 pre-incubated with equimolar concentrations of DNA shows

reduced binding to N₈CRL4^{CRBN} in TR-FRET. Titration of unlabelled IKZF1^{83-196/239-255} (0.02–10 μ M) pre-incubated with control, consensus or no DNA. Unlabelled IKZF1 was added to N₈CRL4^{CRBN}-pomalidomide-
IKZF1 complexes pre-assembled by mixing N₈CRL4^{CRBN} at 250 nM, IKZF1^{1-196/239-255} at 80 nM and pomalidomide at 5 μ M. Addition of buffer to N₈CRL4^{CRBN}-pomalidomide-
IKZF1 (red curve; maximum TR-FRET signal) and N₈CRL4^{CRBN}, IKZF1 without pomalidomide (blue curve; minimum TR-FRET signal). Data are biological replicates presented as means \pm s.d. ($n = 3$).



Extended Data Figure 9 | *In vitro* ubiquitination of CK1 α and identification of target lysines. **a**, *In vitro* ubiquitination of CK1 α by _{N8}CRL4^{CRBN} in the presence different IMiDs at 10 μ M ([IMiD] \gg K_D). **b**, Model of CRL4^{CRBN} bound to CK1 α . Ubiquitin target lysines of CK1 α are shown as spheres and represent sites identified by mass spectrometry following *in vitro* (yellow) or *in vivo*⁸ (blue) ubiquitination. Lys62, 65

and 179 are close to the CUL4 orientation, whereas Lys8 and 225 can be reached in CUL4' orientation. **c**, TR-FRET of _{N8}CRL4^{CRBN}-CK1 α complex formation in the presence of lenalidomide or 2'-deoxyuridine at 5 μ M. 2'-deoxyuridine does not mediate interaction between _{N8}CRL4^{CRBN} and CK1 α . Data are biological replicates presented as means \pm s.d. ($n = 3$).

Extended Data Table 1 | Data collection and refinement statistics

	DDB1 ^{ΔBPB} -CRBN-lenalidomide-CK1α
Data collection	
Space group	<i>P</i> 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	87.97, 109.93, 112.40
α , β , γ (°)	106.02, 93.19, 101.63
Resolution (Å)	65–2.45 (2.58–2.45) *
<i>R</i> _{merge} (%)	7.8 (158)
CC _{1/2} (%)	99.8 (44.7)
<i>I</i> / σ <i>I</i>	8.7 (0.8)
Completeness (%)	97.3 (96.5)
Redundancy	3.6 (3.7)
Refinement	
Resolution (Å)	65–2.45
No. reflections	140745
<i>R</i> _{work} / <i>R</i> _{free}	18.07 / 21.02
No. atoms	
Protein	23199
Ligand/ion	40
Water	305
B-factors	
Protein	97.1
Ligand/ion	78.3
Water	70.2
R.m.s deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.18

*Highest resolution shell is shown in parenthesis.

Structural basis of cohesin cleavage by separase

Zhonghui Lin^{1,2}, Xuelian Luo^{2,3} & Hongtao Yu^{1,2}

Accurate chromosome segregation requires timely dissolution of chromosome cohesion after chromosomes are properly attached to the mitotic spindle. Separase is absolutely essential for cohesion dissolution in organisms from yeast to man^{1,2}. It cleaves the kleisin subunit of cohesin and opens the cohesin ring to allow chromosome segregation. Cohesin cleavage is spatiotemporally controlled by separase-associated regulatory proteins, including the inhibitory chaperone securin^{3–6}, and by phosphorylation of both the enzyme and substrates^{7–12}. Dysregulation of this process causes chromosome missegregation and aneuploidy, contributing to cancer and birth defects. Despite its essential functions, atomic structures of separase have not been determined. Here we report crystal structures of the separase protease domain from the thermophilic fungus *Chaetomium thermophilum*, alone or covalently bound to unphosphorylated and phosphorylated inhibitory peptides derived from a cohesin cleavage site. These structures reveal how separase recognizes cohesin and how cohesin

phosphorylation by polo-like kinase 1 (Plk1) enhances cleavage. Consistent with a previous cellular study¹³, mutating two securin residues in a conserved motif that partly matches the separase cleavage consensus converts securin from a separase inhibitor to a substrate. Our study establishes atomic mechanisms of substrate cleavage by separase and suggests competitive inhibition by securin.

Separase belongs to the clan CD family of cysteine proteases which includes caspases¹. It contains a large amino (N)-terminal armadillo (ARM) repeat domain and a highly conserved carboxy (C)-terminal separase protease domain (SPD) that consists of a pseudo-protease domain (PPD) and an active protease domain (APD)¹⁴ (Fig. 1a and Extended Data Fig. 1). Cohesin forms an asymmetric ring to topologically entrap chromosomes (Fig. 1a)^{15,16}. Separase cleaves the kleisin subunit to open the cohesin ring and trigger chromosome segregation. It also cleaves other substrates to regulate anaphase spindle elongation and centriole duplication^{17,18}. High-resolution structures of separase have not been determined more than a decade

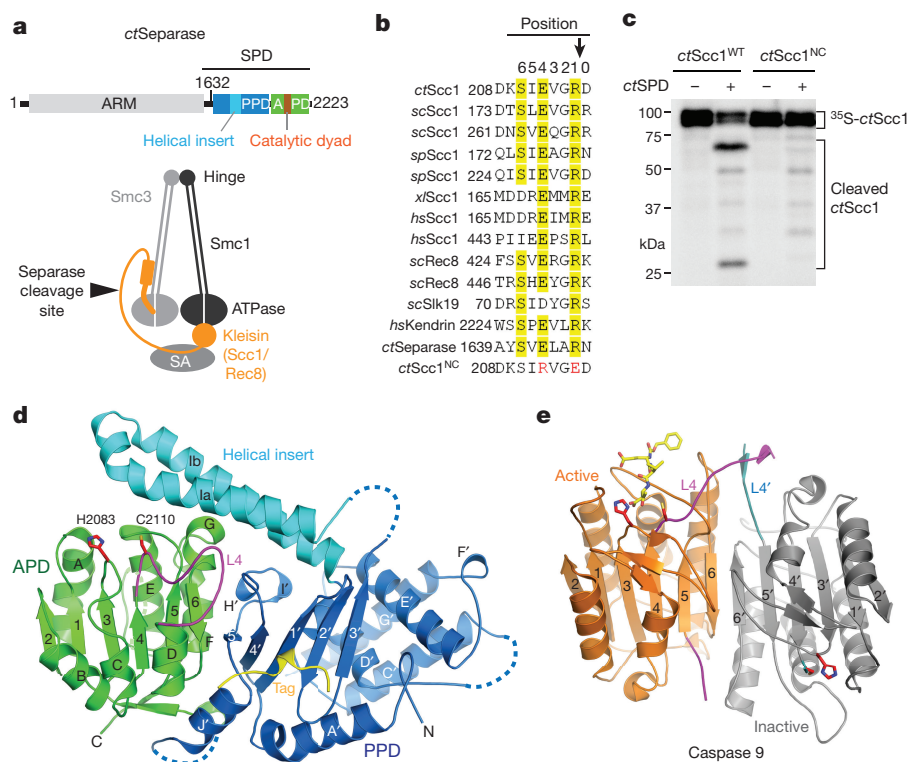


Figure 1 | Structure of *ctSPD*. **a**, Domains and motifs of separase from *C. thermophilum* (top) and schematic drawing of cohesin (bottom). **b**, Sequence alignment of the cleavage sites of separase substrates; *sc*, *Saccharomyces cerevisiae*; *sp*, *Schizosaccharomyces pombe*; *xl*, *Xenopus laevis*; *hs*, *Homo sapiens*. **c**, Autoradiograph of the *ctSPD* cleavage assay with ³⁵S-*ctScc1* wild type (WT) or non-cleavable mutant (NC) as

substrates. For gel source data, see Supplementary Fig. 1. **d**, Cartoon of the crystal structure of *ctSPD*. L4 is coloured magenta. Loops with no visible electron densities are indicated by dashed lines. **e**, Cartoon of caspase 9 (Protein Data Bank accession number 1JXQ), with the bound inhibitor shown as yellow sticks.

¹Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 6001 Forest Park Road, Dallas, Texas 75390, USA. ²Department of Pharmacology, University of Texas Southwestern Medical Center, 6001 Forest Park Road, Dallas, Texas 75390, USA. ³Department of Biophysics, University of Texas Southwestern Medical Center, 6001 Forest Park Road, Dallas, Texas 75390, USA.

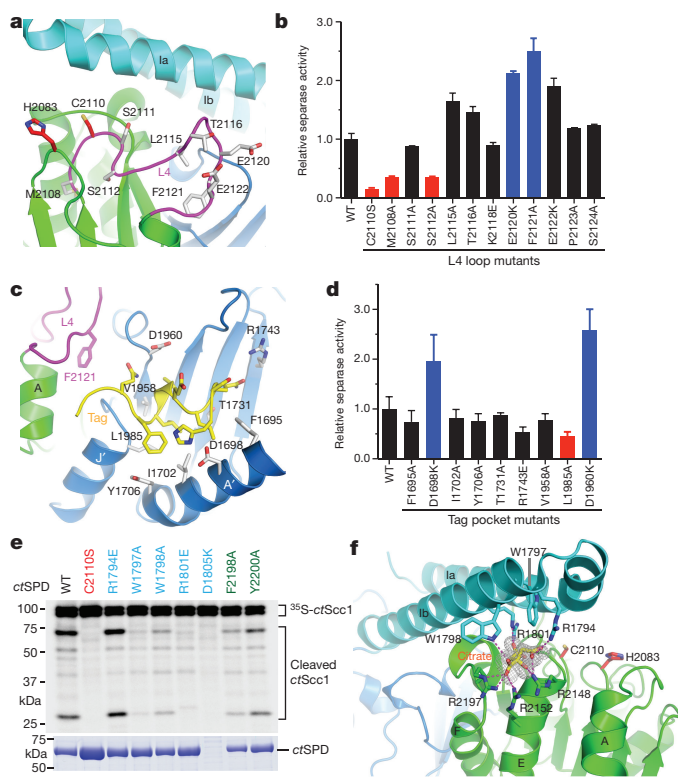


Figure 2 | Contributions of the L4 loop and the helical insert to the activity of *ct*SPD. **a**, Cartoon of *ct*SPD with the catalytic dyad and L4 loop residues shown. **b**, **d**, Quantification of the protease activity of *ct*SPD WT and mutants (mean \pm s.d., $n = 3$ independent experiments). Mutants with activities greater or less than twofold that of WT are in blue and red, respectively. **c**, Interactions between the N-terminal tag and a surface pocket of *ct*SPD. **e**, Autoradiograph of the ^{35}S -*ct*Scc1 cleavage assay by *ct*SPD WT or mutants. Bottom: Coomassie-stained gel of *ct*SPD proteins. **f**, A conserved basic pocket in *ct*SPD, with the $2F_o - F_c$ map of the bound citrate shown at 2.0σ .

since its discovery, hindering our understanding of its mechanism and regulation.

We found that SPD of *C. thermophilum* (*ct*) separase could be expressed in large quantities in bacteria without securin (Extended Data Fig. 2a). Recombinant *ct*SPD, but not the C2110S mutant, cleaved *ct*Scc1 to produce two major fragments (Extended Data Fig. 2b). Separase is known to cleave after the EXXR (X, any residue) consensus motif². Charge-reversal mutation of the $^{212}\text{EVGR}^{215}$ motif in *ct*Scc1 reduced cleavage by separase (Fig. 1b, c). An acyloxymethyl ketone (AMK)-containing peptide inhibitor derived from this cleavage site blocked *ct*Scc1 cleavage in a dose-dependent manner (Extended Data Fig. 2c, d), and retarded the gel mobility of *ct*SPD^{WT}, but not *ct*SPD^{C2110S}, consistent with covalent inhibition (Extended Data Fig. 2e). Similar to separases from other species⁶, longer constructs of *ct*SPD containing an N-terminal extension underwent autocleavage at the $^{1643}\text{ELAR}^{1646}$ site (Fig. 1b and Extended Data Fig. 2f). Thus, recombinant *ct*SPD was active.

We determined the crystal structure of *ct*SPD (Fig. 1d and Extended Data Table 1). It forms one globular domain with two sub-domains—the PPD and the APD—that pack against each other. APD has an overall fold similar to that of caspases (Fig. 1d, e and Extended Data Figs 3 and 4a). PPD also has a mixed α/β fold, but its central β -sheet has a topology different from that of caspases. One edge of this central sheet of PPD forms an edge-on interaction with that of APD, whereas the other edge is capped by a helical domain in PPD. A prominent helical insert of PPD forms a long coiled-coil and packs against APD.

The catalytic dyad H2083 and C2110 are located in loops L3 and L4 of APD (Fig. 1d and Extended Data Fig. 3a). An important mechanism

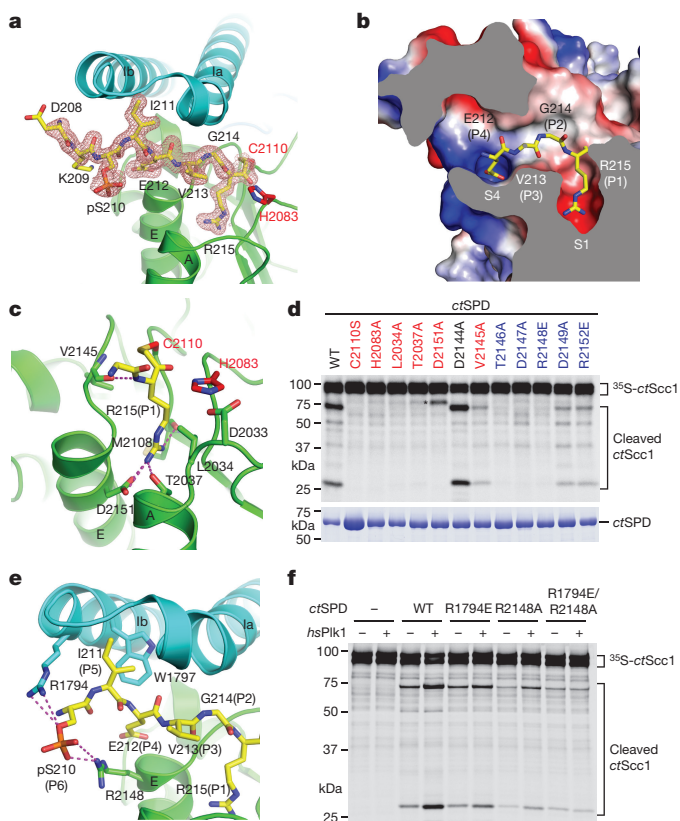


Figure 3 | Structural basis of Scc1 cleavage by separase. **a**, Cartoon of *ct*SPD bound to pAMK (shown as sticks overlaid with the $2F_o - F_c$ map at 1.0σ). **b**, Cross-sectional view of the surface drawing of *ct*SPD-*ct*Scc1 coloured with its electrostatic potential (blue, positive; red, negative; white, neutral). The bound Scc1 peptide is shown as sticks. **c**, The S1 pocket of *ct*SPD that recognizes the P1 arginine. Dashed lines indicate hydrogen bonds or favourable electrostatic interactions. **d**, Autoradiograph of the cleavage reaction of *ct*SPD WT and mutants with ^{35}S -*ct*Scc1 as substrate. Active-site/S1 mutants are labelled red; S4 mutants are labelled blue. Bottom: Coomassie-stained gel of the *ct*SPD proteins. Asterisk marks an aberrant cleavage product of D2151A. **e**, The pS210-binding site. **f**, Autoradiograph of the cleavage reactions of *ct*SPD WT and mutants with ^{35}S -*ct*Scc1 as substrate, with or without a prior incubation with *hs*Plk1.

of pro-caspase activation is the reorganization of L4, which can be achieved through homo-dimerization, cleavage of an internal linker, or both^{19–22}. The geometry of the catalytic dyad and the extended conformation of L4 in *ct*SPD are similar to those in active caspase 9 (Fig. 1d, e), consistent with *ct*SPD being an active enzyme. Thus, separase activation does not require proteolytic cleavage of L4. Consistent with the importance of the L4 loop, mutations of two residues adjacent to C2110, M2108 and S2112, reduced the activity of *ct*SPD (Fig. 2a, b and Extended Data Fig. 4b). In contrast, mutations of L4 residues distal to C2110, including E2120 and F2121, enhanced the activity of *ct*SPD.

A segment of the N-terminal tag of recombinant *ct*SPD binds to a conserved surface pocket in PPD adjacent to L4 (Figs 1d, 2c and Extended Data Fig. 4c). Although this tag is not required for the activity of *ct*SPD, mutations targeting residues in the tag-binding pocket altered the activity of *ct*SPD containing the tag (Fig. 2d and Extended Data Fig. 4d). Similar to mutations of the distal L4 residues, the D1698K and D1960K mutations enhanced the activity of *ct*SPD. We propose that securin or other regions of separase may bind to this tag-binding site, alter the conformation of L4, and affect the protease activity of separase. Even without bona fide ligands, binding of an artificial tag to this site can regulate the protease activity of *ct*SPD in a subtle way.

Unlike active caspase 9, which forms a homodimer¹⁹, separase contains an internal PPD in the same polypeptide chain that packs

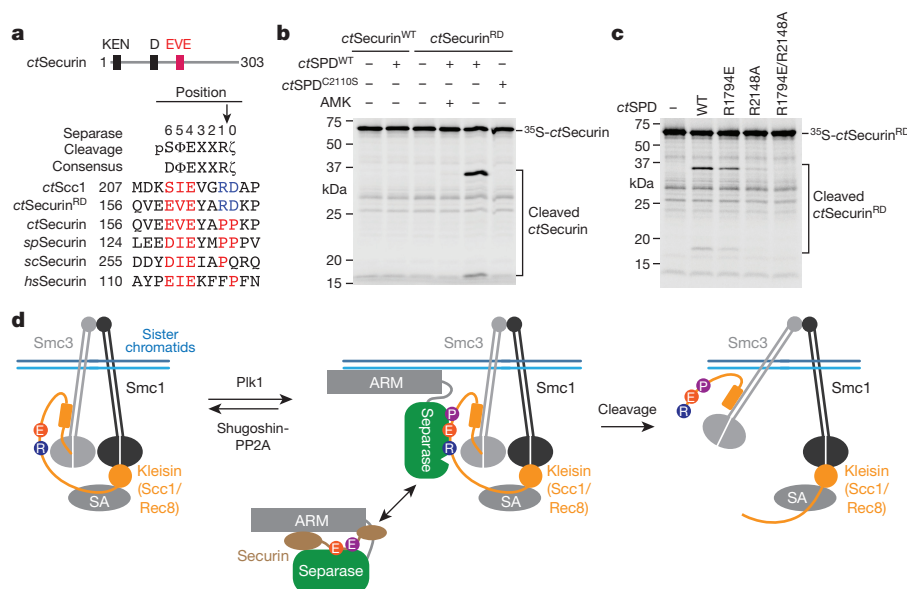


Figure 4 | Securin as a pseudo-substrate of separase. **a**, Sequence alignment of the EVE motif of securin, with the separase cleavage consensus shown above. The KEN and destruction boxes (D) are indicated. Φ/ζ , hydrophobic/hydrophilic residues. **b**, Autoradiograph of the cleavage reactions of ^{35}S -ctSecurin WT or P164R/P165D (RD) by ctSPD WT or C2110S with or without the AMK inhibitor.

c, Autoradiograph of the cleavage reactions of ^{35}S -ctSecurinRD by the indicated ctSPD proteins. **d**, Model depicting specificity determinants, phospho-regulation, and securin inhibition of separase-dependent cohesin cleavage. SA, stromal antigen. Cohesin cleavage by separase can be stimulated by DNA³⁰. The ARM domain of separase might contact DNA.

against and stabilizes its APD. In particular, the helical insert of PPD makes extensive contacts with APD and bridges the two sub-domains (Extended Data Fig. 5a, b). Deletion of the helical insert or mutations of key residues at the helical insert–APD interface, including D1805 and W2143, abolished the expression of soluble ctSPD in bacteria (Fig. 2e and Extended Data Fig. 5a–c). Several helical-insert residues, including C1782 and H1783, are located close to the active site (Extended Data Fig. 5a). Mutations of these residues did not affect the solubility of ctSPD, but reduced the protease activity (Extended Data Fig. 5d). Moreover, residues from the tip of the helical insert, along with residues from APD, form a basic pocket that binds a citrate molecule (Fig. 2f). Mutations of these conserved residues, with the exception of R1794E, diminished separase activity (Fig. 2e and Extended Data Fig. 1). Therefore, the helical insert is critical for both the structure integrity and activity of separase.

Phosphorylation of Scc1 by Plk1 enhances Scc1 cleavage by separase^{10,11}. This cleavage-enhancing phosphorylation is opposed by the shugoshin–PP2A complex bound to cohesin^{23–27}. Incubation of ctScc1, but not ctScc1 S210A, with human Plk1 (hsPlk1) enhanced the cleavage of ctScc1 by ctSPD (Extended Data Fig. 6a). Addition of the hsPlk1 inhibitor BI2536 blocked this enhancement. The phospho-mimicking S210E mutation stimulated ctScc1 cleavage by separase (Extended Data Fig. 6b). Thus, Plk1-dependent phosphorylation of ctScc1 at S210 enhances Scc1 cleavage by separase.

We next determined the crystal structures of ctSPD bound to unphosphorylated (AMK) or phospho-S210-containing (pAMK) inhibitors (Extended Data Table 1). The overall structure of ctSPD–AMK and ctSPD–pAMK complexes is virtually identical to that of free ctSPD, indicating that substrate binding does not induce notable conformational changes. Only the C-terminal²¹²EVGR²¹⁵ segment of the unphosphorylated AMK inhibitor was visible (Extended Data Fig. 6c), whereas all residues of the pAMK inhibitor had clearly defined electron density (Fig. 3a). In both structures, the active-site cysteine C2110 is covalently linked to ctScc1 R215 at the P1 position. Consistent with its role in stabilizing the oxyanion during catalysis, ctSPD H2083 of the catalytic dyad is located close to carbonyl group of ctScc1 R215. R215 forms a salt bridge with ctSPD D2151 at the base of a deep, acidic S1 pocket (Fig. 3b, c). ctScc1 E212 at P4 inserts into the aforementioned

citrate-binding pocket, forming favourable electrostatic and hydrogen bonding interactions (Fig. 3b and Extended Data Fig. 6d). ctScc1 V213 and G214 form minimal contacts with ctSPD. Mutations of residues lining the S1 and S4 pockets in ctSPD greatly diminished separase activity (Fig. 3d).

The S1 pocket mutant D2151A of ctSPD cleaved ctScc1 at a different site (Fig. 3d). The ctScc1 E180K mutation abolished this aberrant cleavage (Extended Data Fig. 6e), indicating that the mutant separase cleaved the ¹⁸⁰ELGM¹⁸³ site. Thus, D2151 not only selects for basic residues but also discriminates against hydrophobic residues at P1. ctSPD charge reversal mutants D2151R and R2152E did not efficiently cleave the complementary charge reversal mutants of ctScc1 (Extended Data Fig. 6f), indicating that other residues in the S1 and S4 pockets contribute to substrate recognition. Because most residues lining the S1 and S4 pockets are conserved among separases in all species (Extended Data Figs 1 and 7), our analyses establish the basis for the EXXR substrate specificity of separase.

Phosphorylation of the substrate does not alter the binding mode of EVGR at P1–P4, but reveals or establishes additional contacts at P5 and P6. I211 at the P5 position packs against W1797 of the helical insert (Fig. 3e). ctScc1 I211A was less efficiently cleaved by ctSPD with or without Plk1 (Extended Data Fig. 8a). Thus, as reported previously²⁸, the hydrophobic residue at P5 contributes to substrate specificity. Phospho-S210 at P6 makes favourable electrostatic interactions with R1794 of the helical insert and R2148 of APD in ctSPD (Fig. 3e and Extended Data Fig. 8b). Single mutation of R1794 or R2148 reduced the stimulation of Scc1 cleavage by Plk1, whereas the double mutation abolished the effect (Fig. 3f and Extended Data Fig. 8c, d). Unlike R2148A, R1794E does not affect the cleavage of unphosphorylated Scc1. Therefore, R1794 specifically serves as a receptor for pS210. R2148 contributes to the recognition of both pS210 and E212. The serine at P6 is conserved in fungal Scc1 and other separase substrates (Fig. 1b). The N-terminal separase cleavage site in vertebrate Scc1 contains a phospho-mimicking, acidic residue at that position. Our structures thus explain the phosphorylation dependency of cohesin cleavage, and further suggest that this phospho-regulation might apply to other separase substrates.

Finally, we probed the mechanism by which securin inhibits separase. Securin blocks substrate access to the active site of separase^{3,6}.

Expectedly, the *ctsecurin-ctseparase* complex was less active in *ctScc1* cleavage, compared with *ctSPD* (Extended Data Fig. 9a, b). A conserved EVE motif in securin matches the separase cleavage consensus at positions P2–P6, but lacks the arginine at P1 and often has a proline at P0 instead of a hydrophilic residue (Fig. 4a). A securin mutant with three residues in this motif mutated was cleaved by separase in fission yeast cells¹³. We thus mutated P164 and P165 in *ctsecurin* to R and D, the matching *ctScc1* residues at P1 and P0. The resulting *ctsecurin*RD mutant was efficiently cleaved by *ctSPD*, and this cleavage was inhibited by the AMK inhibitor (Fig. 4b). Mutating the phosphoserine-binding residues in *ctSPD* or E159 in *ctsecurin*RD reduced cleavage (Fig. 4c and Extended Data Fig. 9c), indicating that this artificial substrate bound at the canonical substrate-binding sites of separase. *ctsecurin* bound tightly to the N-terminal ARM domain of *ctseparase* (Extended Data Fig. 9d). A synthetic EVE-containing securin peptide did not inhibit *ctSPD* (Extended Data Fig. 9e). We propose that securin acts as a pseudo-substrate to competitively block substrate binding to separase (Fig. 4d). Securin binding to the ARM domain of separase provides the necessary avidity for securin to outcompete authentic substrates for access to the active site. Securin is not cleaved because of incompatible residues at the site of cleavage.

As a crucial protease that triggers chromosome segregation, separase is a potential oncoprotein²⁹. Because of the conserved principles of substrate recognition (Extended Data Fig. 7), our structure of an active fungal separase can guide the rational design of chemical inhibitors of human separase, which may have therapeutic potential.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 November 2015; accepted 5 February 2016.

Published online 30 March 2016.

- Uhlmann, F., Wernic, D., Poupert, M. A., Koonin, E. V. & Nasmyth, K. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell* **103**, 375–386 (2000).
- Hauf, S., Waizenegger, I. C. & Peters, J. M. Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. *Science* **293**, 1320–1323 (2001).
- Hornig, N. C., Knowles, P. P., McDonald, N. Q. & Uhlmann, F. The dual mechanism of separase regulation by securin. *Curr. Biol.* **12**, 973–982 (2002).
- Ciosk, R. *et al.* An ESP1/PDS1 complex regulates loss of sister chromatid cohesion at the metaphase to anaphase transition in yeast. *Cell* **93**, 1067–1076 (1998).
- Zou, H., McGarry, T. J., Bernal, T. & Kirschner, M. W. Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* **285**, 418–422 (1999).
- Waizenegger, I., Giménez-Abián, J. F., Wernic, D. & Peters, J. M. Regulation of human separase by securin binding and autocleavage. *Curr. Biol.* **12**, 1368–1378 (2002).
- Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P. & Kirschner, M. W. Dual inhibition of sister chromatid separation at metaphase. *Cell* **107**, 715–726 (2001).
- Gorr, I. H., Boos, D. & Stemmann, O. Mutual inhibition of separase and Cdk1 by two-step complex formation. *Mol. Cell* **19**, 135–141 (2005).
- Hellmuth, S. *et al.* Human chromosome segregation involves multi-layered regulation of separase by the peptidyl-prolyl-isomerase Pin1. *Mol. Cell* **58**, 495–506 (2015).
- Alexandru, G., Uhlmann, F., Mechtler, K., Poupert, M. A. & Nasmyth, K. Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459–472 (2001).
- Hauf, S. *et al.* Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol.* **3**, e69 (2005).
- Katis, V. L. *et al.* Rec8 phosphorylation by casein kinase 1 and Cdc7-Dbp4 kinase regulates cohesin cleavage by separase during meiosis. *Dev. Cell* **18**, 397–409 (2010).

- Nagao, K. & Yanagida, M. Securin can have a separase cleavage site by substitution mutations in the domain required for stabilization and inhibition of separase. *Genes Cells* **11**, 247–260 (2006).
- Viadiu, H., Stemmann, O., Kirschner, M. W. & Walz, T. Domain structure of separase and its binding to securin as determined by EM. *Nature Struct. Mol. Biol.* **12**, 552–553 (2005).
- Gligoris, T. G. *et al.* Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science* **346**, 963–967 (2014).
- Huis in 't Veld, P. J. *et al.* Characterization of a DNA exit gate in the human cohesin ring. *Science* **346**, 968–972 (2014).
- Sullivan, M., Lehane, C. & Uhlmann, F. Orchestrating anaphase and mitotic exit: separase cleavage and localization of Slk19. *Nature Cell Biol.* **3**, 771–777 (2001).
- Matsuo, K. *et al.* Kendrin is a novel substrate for separase involved in the licensing of centriole duplication. *Curr. Biol.* **22**, 915–921 (2012).
- Renatus, M., Stennicke, H. R., Scott, F. L., Liddington, R. C. & Salvesen, G. S. Dimer formation drives the activation of the cell death protease caspase 9. *Proc. Natl Acad. Sci. USA* **98**, 14250–14255 (2001).
- Chai, J. *et al.* Structural basis of caspase-7 inhibition by XIAP. *Cell* **104**, 769–780 (2001).
- Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T. & Alnemri, E. S. Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. *Mol. Cell* **1**, 949–957 (1998).
- Shi, Y. Caspase activation: revisiting the induced proximity model. *Cell* **117**, 855–858 (2004).
- Kitajima, T. S. *et al.* Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441**, 46–52 (2006).
- Riedel, C. G. *et al.* Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* **441**, 53–61 (2006).
- Tang, Z. *et al.* PP2A is required for centromeric localization of Sgo1 and proper chromosome segregation. *Dev. Cell* **10**, 575–585 (2006).
- Ishiguro, T., Tanaka, K., Sakuno, T. & Watanabe, Y. Shugoshin-PP2A counteracts casein-kinase-1-dependent cleavage of Rec8 by separase. *Nature Cell Biol.* **12**, 500–506 (2010).
- Liu, H., Rankin, S. & Yu, H. Phosphorylation-enabled binding of SGO1–PP2A to cohesin protects sororin and centromeric cohesion during mitosis. *Nature Cell Biol.* **15**, 40–49 (2013).
- Sullivan, M., Hornig, N. C., Porstmann, T. & Uhlmann, F. Studies on substrate recognition by the budding yeast separase. *J. Biol. Chem.* **279**, 1191–1196 (2004).
- Zhang, N. *et al.* Overexpression of separase induces aneuploidy and mammary tumorigenesis. *Proc. Natl Acad. Sci. USA* **105**, 13033–13038 (2008).
- Sun, Y. *et al.* Separase is recruited to mitotic chromosomes to dissolve sister chromatid cohesion in a DNA-dependent manner. *Cell* **137**, 123–132 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Rosenbaum for the *C. thermophilum* complementary DNA, H. Ball for peptide synthesis, and D. Tomchick and Z. Chen for assistance with data collection. Diffraction data of the selenomethionine separase were collected at the Advanced Light Source at Lawrence Berkeley National Laboratory with the help of its staff. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract number DE-AC02-05CH11231. Results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by UChicago Argonne, LLC, for the US Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. This work is supported by the Cancer Prevention and Research Institute of Texas (RP110465-P3 to H.Y.), the National Institutes of Health (GM107415 to X.L.), and the Welch Foundation (I-1441 to H.Y.). H.Y. is an investigator with the Howard Hughes Medical Institute.

Author Contributions Z.L. performed all experiments in this study with advice from H.Y. X.L. provided assistance with structure refinement. Z.L. and H.Y. wrote the paper.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 5FBY, 5FC3, and 5FC2. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Y. (hongtao.yu@utsouthwestern.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Expression and purification of *ctSPD*. The *ctseparase* cDNA (GenBank identity 18261092) was synthesized at GenScript USA. For the expression of the *ctSPD*, the cDNA fragment of *ctSPD*^{1632–2223} was subcloned into a modified pET bacterial expression vector. The pET-*ctSPD* vector encoded *ctSPD*^{1632–2223} with an N-terminal His₆ tag of the following sequence: MGSSHHHHHHSQLEVLFGQPLGSGRP. The pET-*ctSPD* vector was transformed into *Escherichia coli* strain BL21 (DE3). Protein expression was induced with isopropylthiogalactoside (IPTG) at 18 °C overnight. The bacteria were harvested and resuspended in the lysis buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol, 1 mM DTT, and 0.05% Triton X-100). After sonication and centrifugation, the supernatant was applied to Ni²⁺-NTA resin (Qiagen). After extensive washing, His₆-*ctSPD* was eluted from the Ni²⁺-NTA column. His₆-*ctSPD* was further purified with a mono Q 5/50 GL anion-exchange column (GE Healthcare) and a Superdex 200 10/300 GL column. The point mutants of *ctSPD* were generated with a QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies). The truncated variants and point mutants of *ctSPD* were expressed and purified similarly. Because *ctSPD*^{1632–2223} underwent autocleavage at the ¹⁶⁴³ELAR¹⁶⁴⁶ site, we generated a non-cleavable ¹⁶⁴³RLAE¹⁶⁴⁶ mutant to prevent autocleavage and increase yield. All *ctSPDs* in this study, except that in Extended Data Fig. 2f, contained the non-cleavable mutation.

The selenomethionine (SeMet)-labelled *ctSPD*^{1663–2223} was produced with the methionine biosynthesis inhibition method³¹. Briefly, bacteria transformed with pET-*ctSPD* cultured overnight were pelleted, washed, and resuspended with M9 minimal media. The bacteria were further incubated at 37 °C until the absorbance at 600 nm reached about 1.0. Methionine biosynthesis was inhibited by the addition of the amino-acid solution containing 50 mg l⁻¹ of Leu/Ile/Val and 100 mg l⁻¹ of Phe/Lys/Thr/SeMet. Protein expression was induced with 0.4 mM IPTG at 18 °C overnight. The SeMet-labelled protein was subsequently purified through the same procedure as described above.

Expression and purification of *ctseparase*–*securin* complex and *ctSPD*^{1501–2223}. The *ctsecurin* cDNA (GenBank identity 18256826) was cloned from a cDNA library of *C. thermophilum*. The cDNAs of *ctseparase* and *ctsecurin* were separately subcloned into a modified pFastBac HT vector (Invitrogen). The final constructs encoded an N-terminal His₆-Strep-tagged *ctseparase* and an N-terminal His₆-tagged *ctsecurin*. Baculoviruses of *ctseparase* (full-length or residues 1–1500) and *ctsecurin* were constructed with the Bac-to-Bac system (Invitrogen) according to the manufacturer's protocols. Sf9 cells were co-infected with *ctseparase* and *ctsecurin* baculoviruses and harvested at 48 h after infection. Cells were resuspended in the lysis buffer containing 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol, 1 mM DTT, and 0.05% Triton X-100, followed by sonication and centrifugation. The supernatant was applied onto a Strep-Tactin Superflow column (Qiagen). After extensive washes with the lysis buffer, the *ctseparase*–*securin* complex was eluted with the elution buffer containing 5 mM d-Desthiobiotin (Sigma-Aldrich), 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol, and 1 mM DTT. The His₆-*ctSPD*^{1501–2223} protein was expressed in Sf9 cells with a similar strategy and purified through a Ni²⁺-NTA column.

Separase activity assay. The *ctScc1* cDNA (GenBank identity 18259702) was synthesized by GenScript USA and was cloned into a modified pCS2 vector with a SP6 promoter. To produce ³⁵S-*ctScc1* or its mutants, the pCS2-*ctScc1* plasmids were added to a TNT Quick Coupled Transcription Translation System (Promega) and incubated in the presence of ³⁵S-methionine at 30 °C for 90 min. (The ³⁵S-*ctScc1* proteins migrated as a doublet on SDS–polyacrylamide gel electrophoresis (SDS–PAGE), possibly owing to proteolysis or internal methionine initiation during the *in vitro* translation reaction.) Then, 2 µl of ³⁵S-*ctScc1* was added to 18 µl of *ctSPD* (~1.5 µM) or *ctseparase*–*securin* protein solution containing 25 mM HEPES (pH 7.5), 75 mM KCl, 5 mM MgCl₂, 1 mM DTT, 15 mM NaF, 1 mM EGTA, 10% glycerol, and 0.05% Triton X-100, and incubated at 30 °C for 60 min. For assays in Fig. 1c and Extended Data Fig. 2b, a higher concentration of *ctSPD* (3.0 µM) was used, resulting in more complete *ctScc1* cleavage. For AMK inhibition assay, *ctSPD* was pre-incubated with the *ctScc1*-AMK inhibitor (synthesized by KareBay Biochem) at room temperature (25 °C) for 30 min and further incubated with ³⁵S-*ctScc1* at 30 °C for 60 min. In the Plk1 stimulation assay, ³⁵S-*ctScc1* was pre-treated with recombinant GST-*hsPlk1*^{T210D} in the kinase buffer for 30 min at 30 °C, in the absence or presence of 10 µM BI2536, and further incubated with *ctSPD* (0.5 µM) for 60 min. The reaction mixtures were separated on SDS–PAGE gels, which were stained, destained, dried, and analysed with Fuji or GE phosphorimagers.

The *ctsecurin* WT and P164R/P165D (*ctsecurin*RD) cDNAs were subcloned into pCS2-Myc vector with an SP6 promoter. The ³⁵S-*ctsecurin*^{WT} and *ctsecurin*RD proteins were produced and assayed as described above. The *ctsecurin*_{153–177} (DPLQVEVEVAPPKPKEMPYESDVF) and *ctsecurin*_{153–177} 3A (DPLQVEAAAYAPPKPKEMPYESDVF) peptides were chemically synthesized and tested for their ability to inhibit the cleavage of *ctScc1* by *ctSPD* as described above.

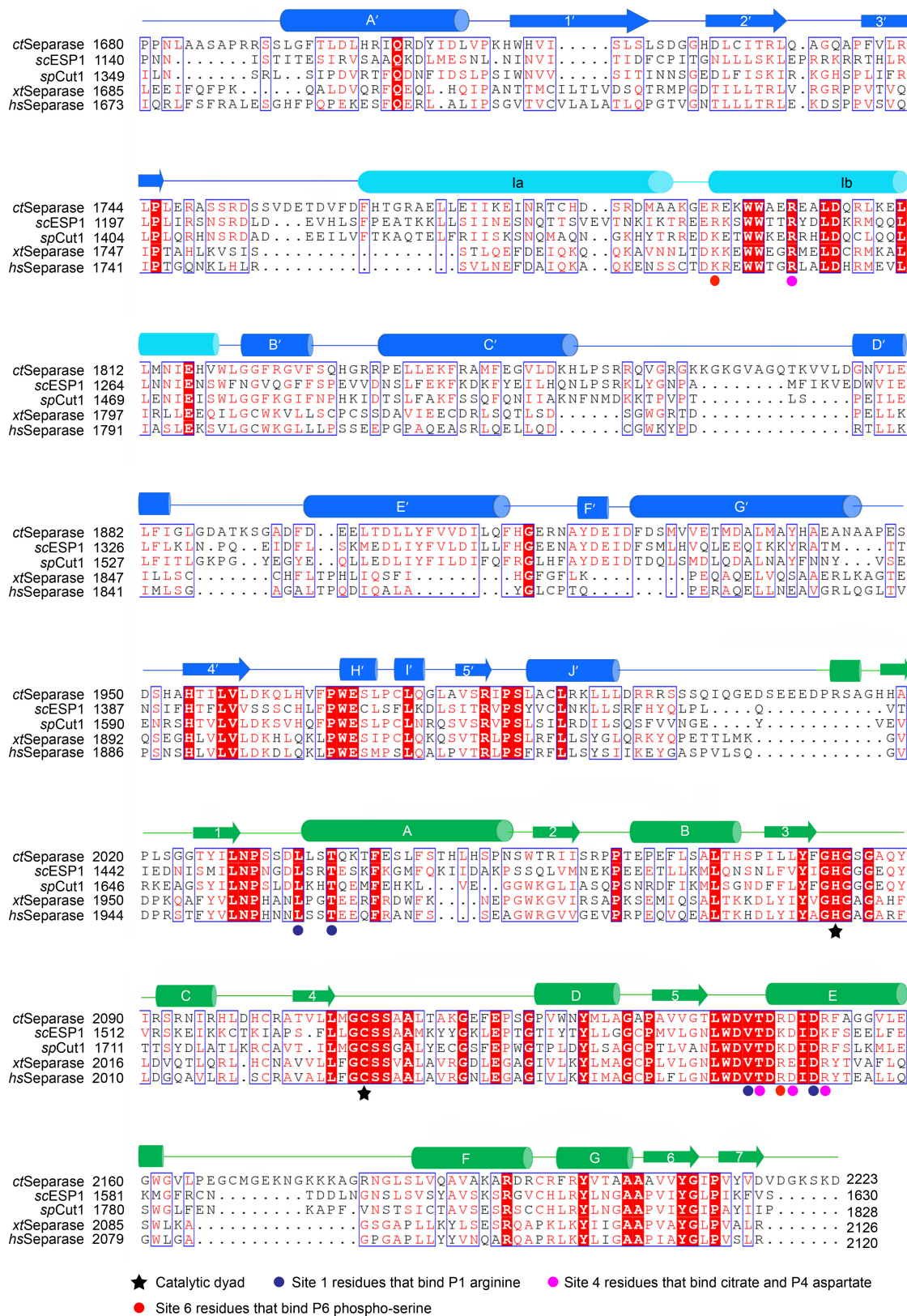
Crystallization and data collection. All crystallization experiments were performed at 20 °C. Initial screens were performed with a Phoenix crystallization robot (Art Robbins Instruments), using the commercially available screening kits from Hampton Research, Qiagen, and Molecular Dimensions. Conditions obtained from the initial screens were optimized using hanging-drop vapour diffusion method. Diffraction-quality crystals were obtained by repeated micro-seeding. All crystals were cryoprotected with a reservoir solution supplemented with 15% glycerol.

Both native and SeMet-labelled *ctSPD*^{1663–2223} crystals were grown by mixing equal volumes of the protein solution (11 mg ml⁻¹) with the precipitant solution containing 0.2 M ammonium citrate tribasic (pH 7.0), 20% PEG3350, and 10 mM DTT. Diffraction data were collected at beamline BL8.2.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at the wavelength of 0.9786 Å at 100 K and processed with HKL3000 (ref. 32).

For crystallization of *ctSPD*^{1632–2223}-AMK and *ctSPD*^{1693–2223}-pAMK complexes, the purified *ctSPD* proteins were mixed with the *ctScc1*-AMK or phospho-*ctScc1*-AMK peptide inhibitors (KareBay Biochem) at a molar ratio of 1:2.5, and incubated overnight at room temperature to form covalent complexes as monitored by SDS–PAGE. The complexes were further purified with a Superdex 200 10/300 GL size-exclusion column in the buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, and 5 mM DTT. Crystals of *ctSPD*^{1632–2223}-AMK were grown by mixing equal volumes of the protein solution (13 mg ml⁻¹) with the precipitant solution containing 0.1 M ammonium citrate tribasic (pH 7.0) and 12% PEG3350. For the crystallization of *ctSPD*^{1693–2223}-pAMK complex, the 11 mg ml⁻¹ protein solution was mixed with an equal volume of the precipitant solution containing 0.2 M KCl, 50 mM HEPES (pH 7.5), 32% pentaerythritol propoxylate (5/4 PO/OH), and 10 mM DTT. Diffraction data for *ctSPD*^{1632–2223}-AMK and *ctSPD*^{1693–2223}-pAMK were collected at beamline 19-ID (SBC-CAT) at the Advanced Photon Source (Argonne National Laboratory) at 100 K at wavelengths of 0.9793 Å and 0.9795 Å, respectively, and processed with HKL3000.

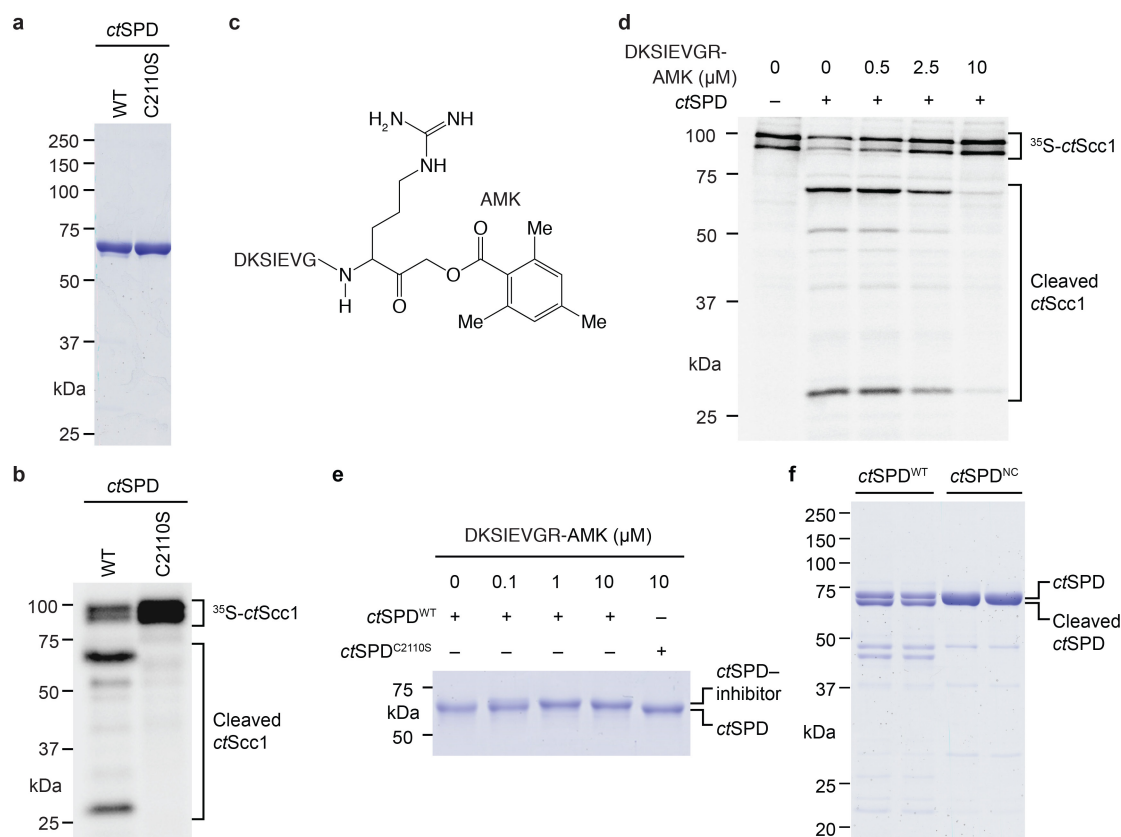
Structure determination and refinement. The crystal of SeMet-labelled *ctSPD*^{1663–2223} diffracted to a minimum Bragg spacing of 2.20 Å and exhibited the symmetry of space group P2₁2₁2₁ with cell dimensions of *a* = 55.67 Å, *b* = 98.79 Å, *c* = 107.76 Å. Phases were obtained from the selenium single-wavelength anomalous diffraction method. With data truncated to 2.5 Å, nine of ten possible selenium sites were located and refined with PHENIX AutoSol³³, resulting in an overall figure of merit of 0.323. The experimental electron density map was used to construct an initial model with automated building with PHENIX AutoBuild. As a result, 414 of total 587 residues were built in the initial model, with *R*_{work} and *R*_{free} of 27.74% and 32.79%, respectively. Iterative model building and refinement were performed with JLigand³⁴, COOT³⁵, and PHENIX. Phases of native *ctSPD*^{1663–2223}, *ctSPD*^{1632–2223}-AMK, and *ctSPD*^{1693–2223}-pAMK were obtained by molecular replacement with Phaser using the SeMet crystal structure as the search model. Data collection and structure refinement statistics are summarized in Extended Data Table 1. Ramachandran statistics (favoured/allowed/outlier (%)) calculated by MolProbity³⁶ for *ctSPD*^{1663–2223}, *ctSPD*^{1632–2223}-AMK, and *ctSPD*^{1693–2223}-pAMK were 98.0/1.6/0.4, 98.1/1.9/0.0, and 98.1/1.7/0.2, respectively. All structural figures were generated with the program PyMOL (<http://www.pymol.org/>) using the same colour and labelling schemes.

- Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
- Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Lebedev, A. A. *et al.* J. Ligand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D* **68**, 431–440 (2012).
- Emsley, P. & Cowtan, K. COOT: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).



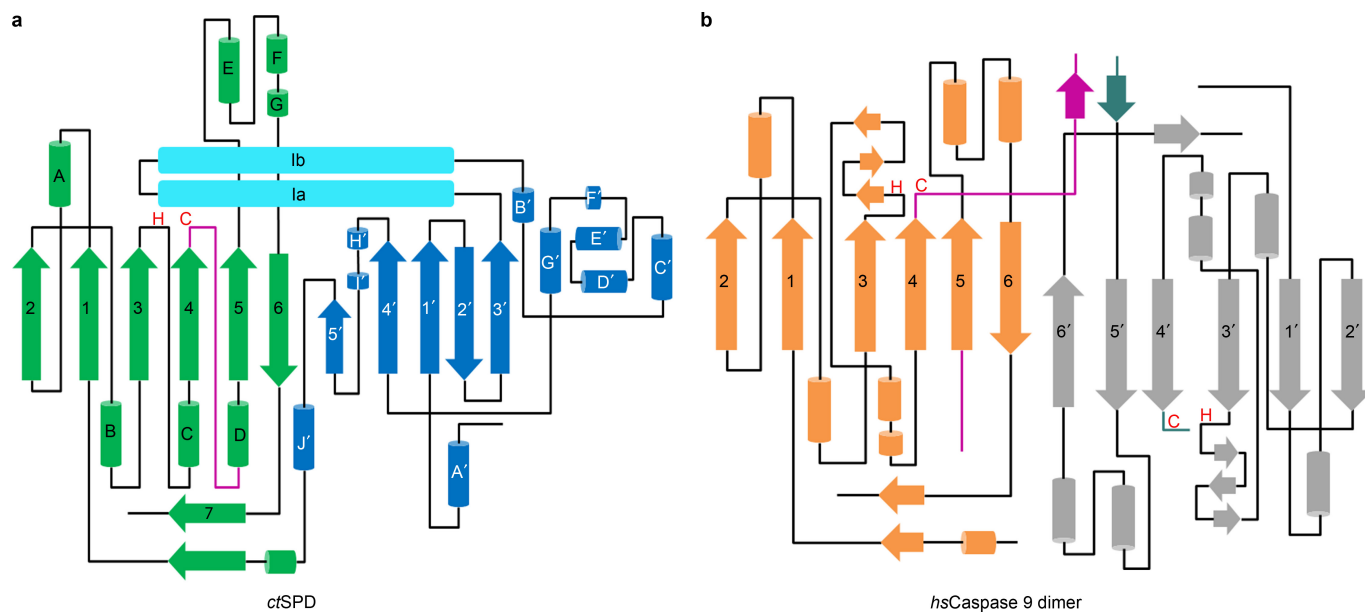
Extended Data Figure 1 | Sequence alignment of the SPDs from multiple species. The alignment is generated using the online ESPrnt 2.0 server. Secondary structural elements of ctSPD are indicated above the sequences, with the same labelling and colour schemes as in Fig. 1d

(PPD, blue; APD, green; the helical insert in PPD, cyan). Abbreviations: *ct*, *Chaetomium thermophilum*; *sc*, *Saccharomyces cerevisiae*; *sp*, *Schizosaccharomyces pombe*; *xt*, *Xenopus tropicalis*; *hs*, *Homo sapiens*.

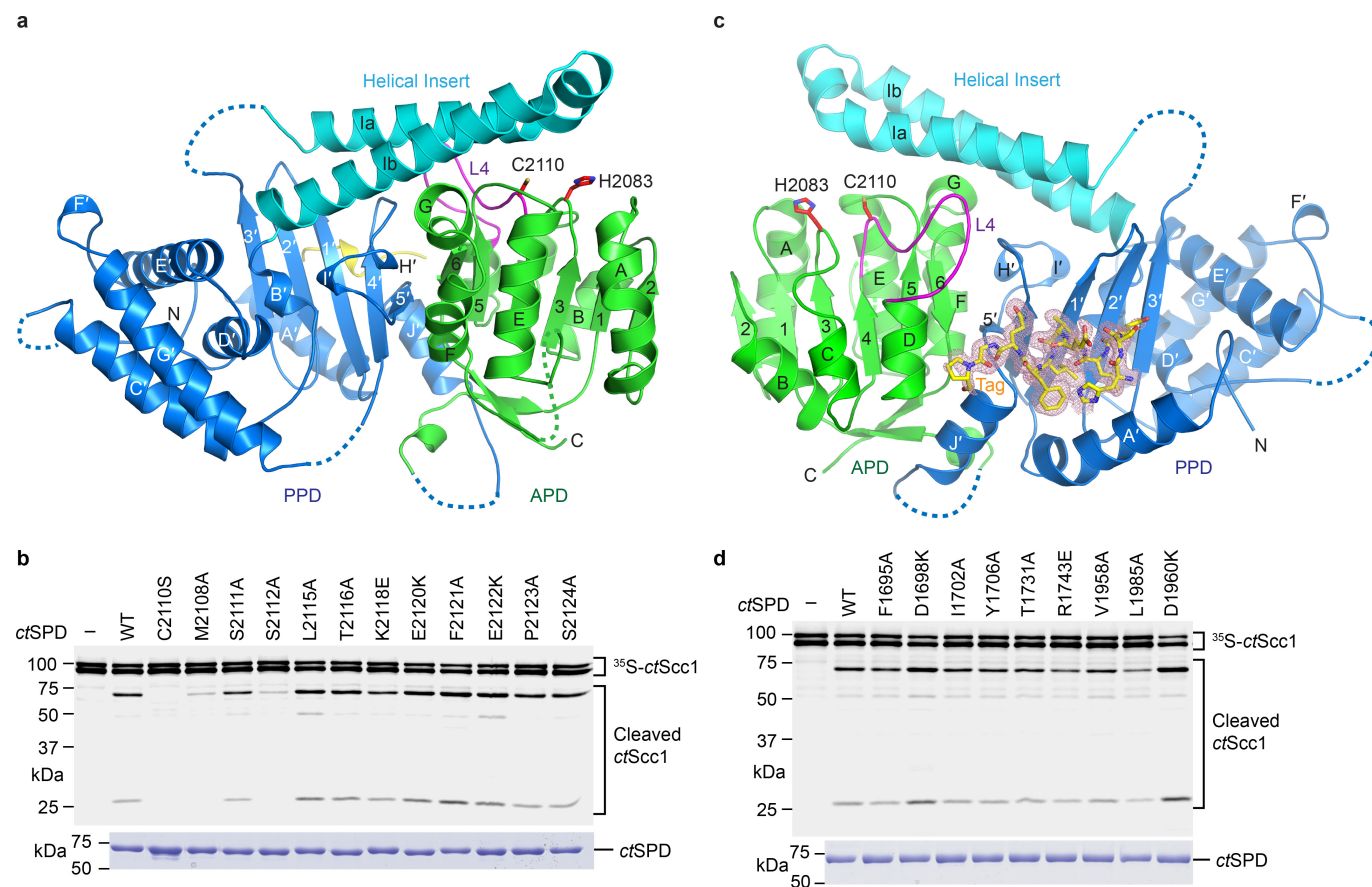


Extended Data Figure 2 | Purification, activity, inhibition, and autocleavage of active *ctSPD*. **a**, Coomassie-stained gel of purified recombinant *ctSPD* wild type (WT) and C2110S. **b**, Autoradiograph of the *ctSPD* cleavage assay with ^{35}S -*ctScc1* as substrate. **c**, Chemical structure of the acyloxymethyl ketone (AMK) inhibitor derived from the *ctScc1* cleavage site. **d**, Autoradiograph of the *ctSPD* cleavage assay with ^{35}S -*ctScc1* as substrate, in the absence or presence of increasing doses of

the AMK inhibitor depicted in **c**. **e**, Coomassie-stained SDS-PAGE gel of purified recombinant *ctSPD* WT or C2110S treated with the indicated doses of the *ctScc1*-AMK peptide inhibitor. The positions of unmodified *ctSPD* and *ctSPD*-inhibitor conjugates are indicated. **f**, Coomassie-stained gel of recombinant *ctSPD*^{1632–2223} WT or non-cleavable (NC) mutant. The *ctSPD*^{NC} mutant contains the E1643R and R1646E mutations. The positions of intact and autocleaved *ctSPD* proteins are indicated.

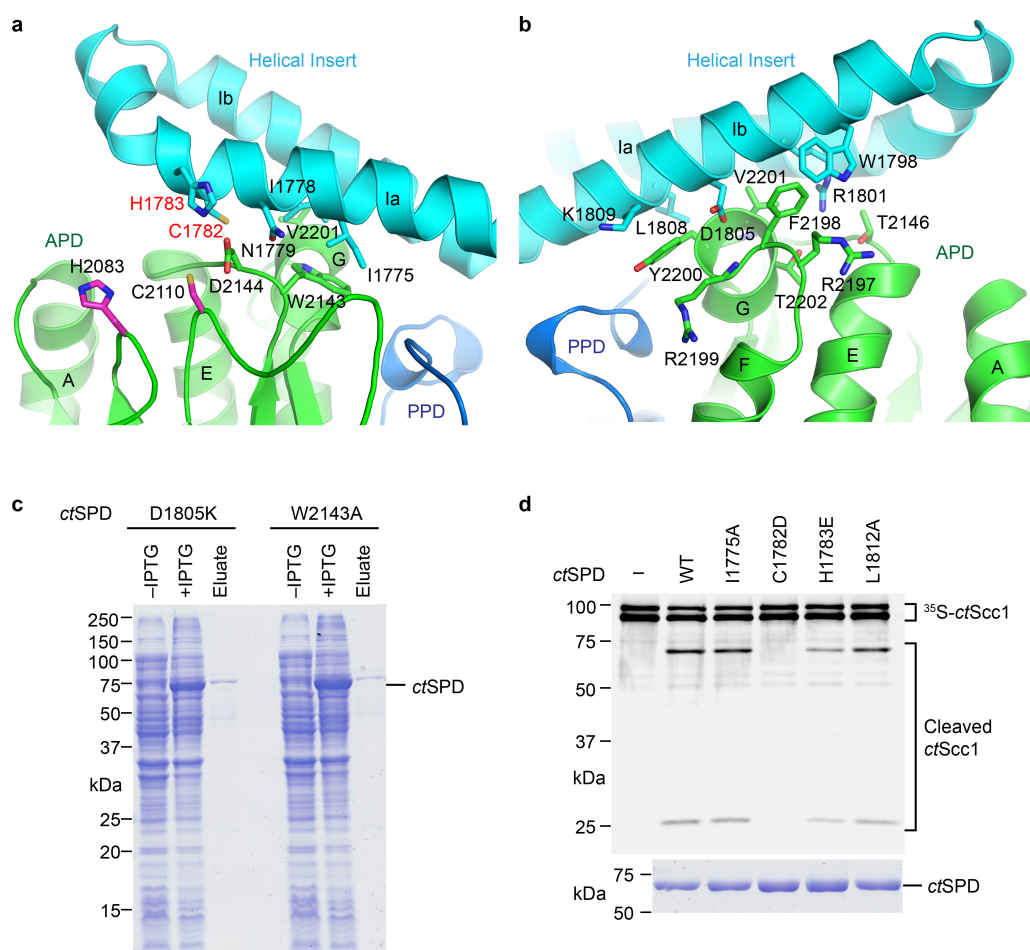


Extended Data Figure 3 | Comparison between the folding topologies of *ctSPD* and the caspase 9 dimer (Protein Data Bank accession number 1JXQ). The labelling and colour schemes are the same as in Fig. 1d, e. H, the catalytic histidine; C, the catalytic cysteine.



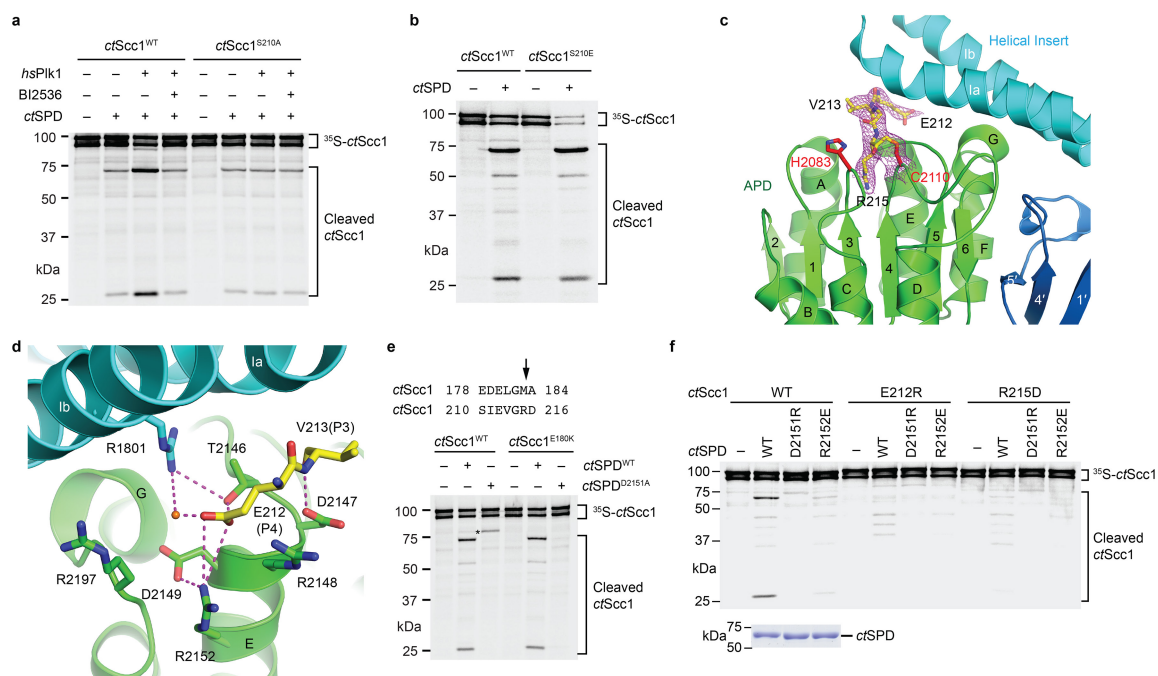
Extended Data Figure 4 | Contributions of the L4 loop and a surface pocket to the protease activity of *ctSPD*. **a**, Cartoon of the crystal structure of *ctSPD*, with the PPD coloured blue, the APD coloured green, the helical insert in PPD coloured cyan, and an N-terminal tag peptide coloured yellow. The N and C termini are indicated. All secondary structure elements are labelled. Loops with no visible electron densities are indicated by dashed lines. Loop 4 (L4) is coloured magenta. H2083 and C2110 of the catalytic dyad are shown as sticks. The orientation of *ctSPD* in this figure is related to that in Fig. 1d by a 180° rotation along the vertical axis. **c**, Representative autoradiograph of the ³⁵S-*ctScc1* cleavage assay by WT *ctSPD* or the indicated mutants. Bottom: Coomassie-stained

gel of *ctSPD* proteins used in the assay. Quantification of the relative protease activities of *ctSPD* WT and mutants is shown in Fig. 2b. The protease activity is defined as the ratio between intensities of the two major *ctScc1* cleavage products and that of the uncleaved *ctScc1*. **c**, Cartoon of the crystal structure of *ctSPD*, in the same orientation as in Fig. 1d. The tag peptide (HSQLEVLFGQP) is shown as sticks, overlaid with its $2F_o - F_c$ electron density map contoured at 1.0σ . **d**, Representative autoradiograph of the ³⁵S-*ctScc1* cleavage assay by WT *ctSPD* or the indicated mutants. Bottom: Coomassie-stained gel of *ctSPD* proteins used in the assay. Quantification of the relative protease activities of *ctSPD* WT and mutants is shown in Fig. 2d.



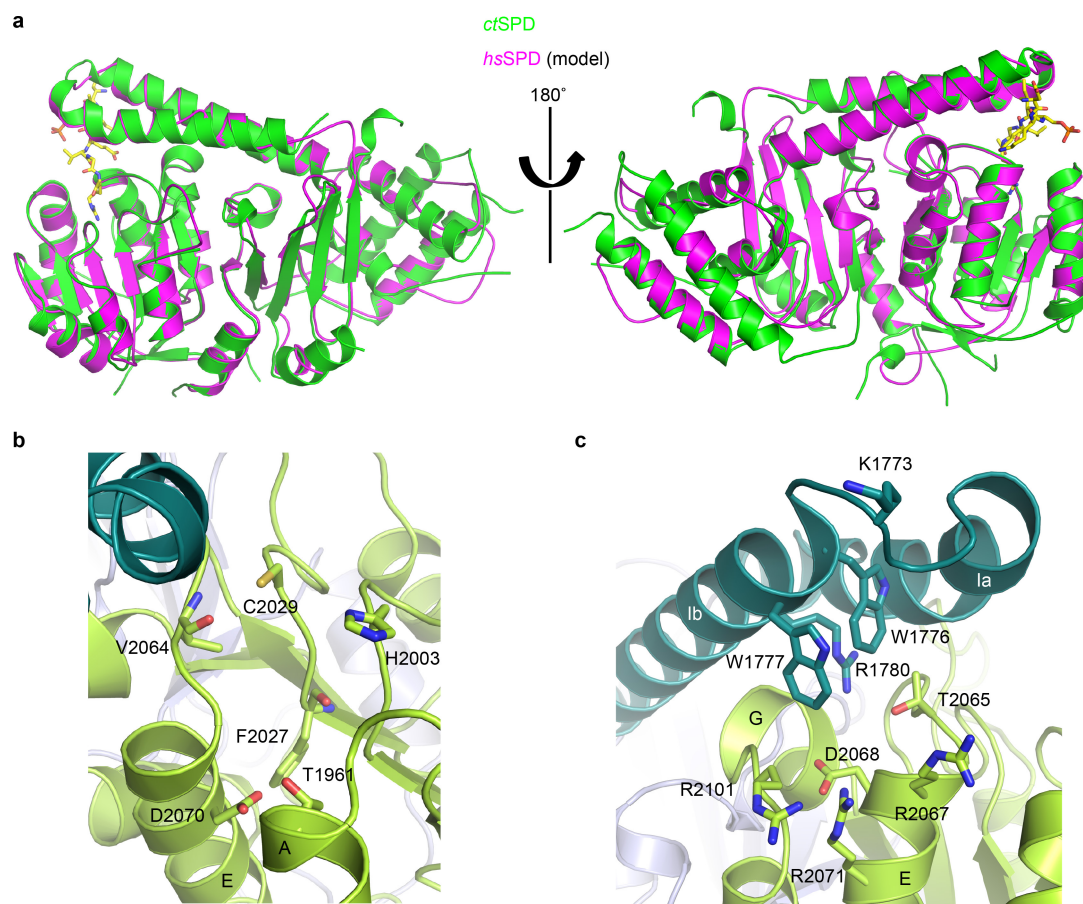
Extended Data Figure 5 | Interactions between the helical insert and the APD. **a, b**, Zoomed-in views of cartoons of *ctSPD* in two orientations that are related by a 180° rotation along the vertical axis. Residues at the interface between the helical insert of the PPD and the APD are shown in sticks and labelled. **c**, Coomassie-stained gel of lysates of bacteria expressing the indicated *ctSPD* mutants and treated without (–) or with

(+) isopropyl β-D-1-thiogalactopyranoside (IPTG) and eluates from Ni²⁺-NTA beads that had been incubated with the IPTG lysates. **d**, Autoradiograph of the ³⁵S-*ctScc1* cleavage assay by WT *ctSPD* or the indicated mutants. Bottom: Coomassie-stained gel of *ctSPD* proteins used in the assay.



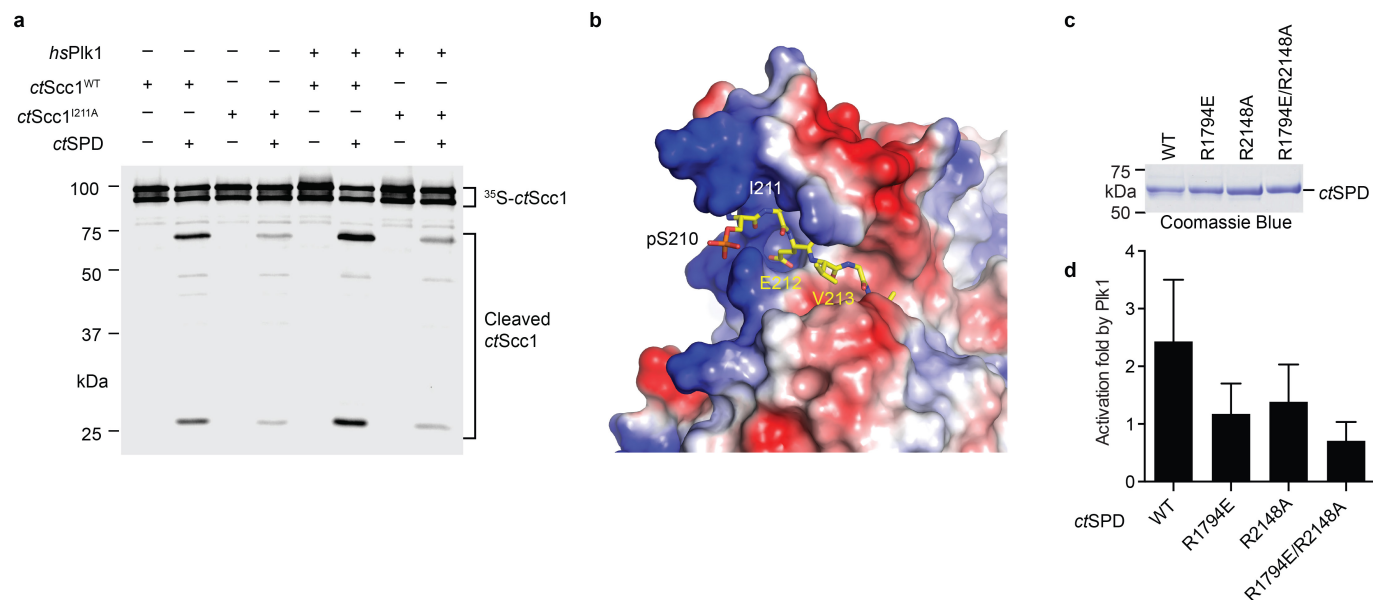
Extended Data Figure 6 | Phospho-regulation and specificity determinants of separase-mediated cohesin cleavage. **a**, Autoradiograph of the *ctSPD* cleavage reactions of ³⁵S-*ctScc1* WT or S210A, treated with or without human (*hs*) Plk1 or its inhibitor BI2536. **b**, Autoradiograph of the *ctSPD* cleavage reactions, with ³⁵S-*ctScc1* WT or the phospho-mimicking S210E as substrates. **c**, Zoomed-in view of the cartoon of *ctSPD* bound covalently to the *ctScc1*-AMK inhibitor. The catalytic dyad residues C2110 and H2083 are shown as red sticks. The covalently bound inhibitor is shown as yellow sticks, overlaid with its $2F_o - F_c$ electron density map contoured at 1.0σ . **d**, Zoomed-in view of the S4 pocket of *ctSPD* that recognizes the P4 glutamate. Dashed lines indicate hydrogen

bonds or favourable electrostatic interactions. The orange sphere indicates a water molecule. **e**, Mapping of the aberrant *ctScc1* cleavage site by *ctSPD* D2151A. Top: sequence alignment of the aberrant site of D2151A and the major site of WT. Bottom: autoradiograph of the cleavage reactions of *ctSPD*^{WT} or *ctSPD*^{D2151A} with the indicated ³⁵S-*ctScc1* proteins as substrates. Asterisk marks the aberrant cleavage product by *ctSPD*^{D2151A}. **f**, Charge reversal mutants of *ctSPD* fail to cleave complementary charge reversal mutants of *ctScc1*. Autoradiograph of the cleavage assay of WT *ctSPD* or the indicated mutants, with ³⁵S-*ctScc1* WT or mutants as substrates. Bottom: Coomassie-stained gel of *ctSPD* proteins used in the assay.



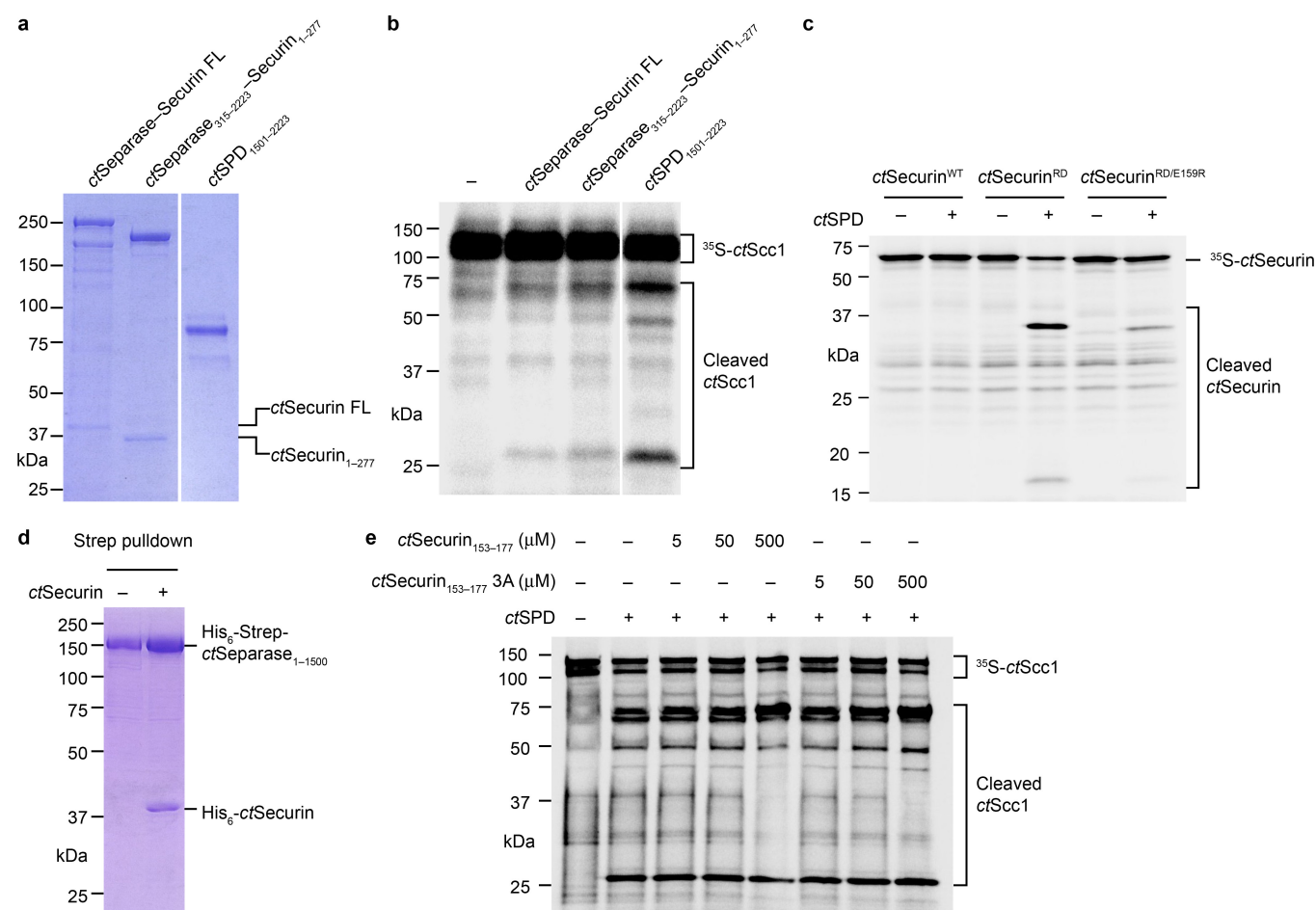
Extended Data Figure 7 | Conservation of substrate-binding residues in human separase. **a**, Two different views of the cartoon of the structure of α SPD–pAMK (green) and a homology model of human (hs) SPD (magenta). The phospho-AMK peptide is shown in sticks. The homology

model of hs SPD was generated with SWISS-MODEL. The coordinates of the model are available upon request. **b**, **c**, Zoomed-in views of the S1 and S4 pockets of the hs SPD model.



Extended Data Figure 8 | Structural basis of phosphorylation-stimulated Scc1 cleavage. **a**, Autoradiograph of the *ctSPD* cleavage reactions of ³⁵S-*ctScc1* WT or I211A, treated with or without *hsPlk1*. **b**, Zoomed-in view of the surface drawing of *ctSPD*–pAMK. The surface is coloured according to the electrostatic potential, with red, blue, and white representing negative, positive, and neutral charges, respectively.

The covalently bound peptide is shown as sticks. **c**, Coomassie-stained gel of the indicated *ctSPD* proteins used in the assays described in Figs 3f and 4c. **d**, Quantification of the fold of Plk1 stimulation in *ctScc1* cleavage by *ctSPD* WT and the indicated mutants as described in Fig. 3f. Error bars, s.d. ($n = 3$ independent experiments).



Extended Data Figure 9 | Interactions between *ctsecurin* and *ctseparase*.

a, Coomassie-stained gel of recombinant *ctseparase*-*ctsecurin* complexes and *ctSPD* expressed in insect cells. FL, full length. **b**, Autoradiograph of the *ctScc1* cleavage reactions by the *ctseparase*-*ctsecurin* complexes and *ctSPD*. **c**, Autoradiograph of the cleavage reactions of ^{35}S -*ctsecurin* WT or mutants with or without *ctSPD*. **d**, Coomassie-stained gel of recombinant

Strep-tagged *ctseparase*₁₋₁₅₀₀ or the *ctseparase*₁₋₁₅₀₀-*ctsecurin* complex bound to Strep-Tactin beads. **e**, Autoradiograph of the *ctScc1* cleavage reactions by *ctSPD*, in the absence or presence of varying concentrations of the *ctsecurin*₁₅₃₋₁₇₇ or *ctsecurin*₁₅₃₋₁₇₇ 3A peptides. The EVE motif is mutated to AAA in the *ctsecurin*₁₅₃₋₁₇₇ 3A peptide.

Extended Data Table 1 | Data collection and refinement statistics

	<i>ctSPD</i> ¹⁶⁶³⁻²²²³	<i>ctSPD</i> ¹⁶³²⁻²²²³ -AMK	<i>ctSPD</i> ¹⁶⁹³⁻²²²³ -pAMK
Data collection			
Space group	P2 ₁ 2 ₁ 2 ₁	P6 ₃ 22	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	55.56, 98.89, 107.75	149.15, 149.15, 115.63	56.35, 85.01, 119.27
α , β , γ (°)	90, 90, 90	90, 90, 120	90, 90, 90
Resolution (Å)	50.00-1.90 (1.93-1.90) *	50.00-3.10 (3.15-3.10)	50.00-1.85 (1.88-1.85)
<i>R</i> _{merge} (%)	7.4 (97.0)	15.5 (100)	13.5 (100)
<i>I</i> / σ <i>I</i>	27.7 (2.5)	14.6 (1.3)	19.8 (1.9)
Completeness (%)	100 (100)	99.9 (98.7)	100 (100)
Redundancy	7.2 (7.3)	9.6 (8.7)	7.5 (5.8)
Refinement			
Resolution (Å)	49.38-1.90	48.82-3.10	42.51-1.85
No. reflections	46925	14228	50139
<i>R</i> _{work} / <i>R</i> _{free}	17.3 / 20.0	19.4 / 25.8	18.0 / 20.7
No. atoms			
Protein	3937	3745	3730
Ligand/ion	13	36	134
Water	394	52	454
<i>B</i> -factors			
Protein	27.4	78.4	26.6
Ligand/ion	32.6	86.0	29.8
Water	37.1	66.3	38.4
R.m.s deviations			
Bond lengths (Å)	0.012	0.010	0.013
Bond angles (°)	1.24	1.13	1.05

Data were collected from one crystal for each structure.

*Highest-resolution shell is shown in parenthesis.

TOOLBOX

THE ROBOTS THAT TAP INTO BRAIN CELLS

Neuroscientists hope to turn the delicate art of eavesdropping on neurons into an automated technique.

ILLUSTRATION BY THE PROJECT TWINS



BY HELEN SHEN

Clamping an electrode to the brain cell of a living animal to record its electrical chatter is a task that demands finesse and patience. Known as ‘whole-cell patch-clamping’, it is reputedly the “finest art in neuroscience”, says neurobiologist Edward Boyden, and one that only a few dozen laboratories around the world specialize in.

But researchers are trying to demystify this art by turning it into a streamlined, automated technique that any laboratory could attempt, using robotics and downloadable source code.

“Patch-clamping provides a unique view

into neural circuits, and it’s a very exciting technique but is really underused,” says neuroscientist Karel Svoboda at the Howard Hughes Medical Institute’s Janelia Research Campus in Ashburn, Virginia. “That’s why automation is a really, really exciting direction.”

On 3 March, Boyden, at the Massachusetts Institute of Technology in Cambridge, and his colleagues published detailed instructions on how to assemble and operate an automated system for whole-cell patch-clamping¹, a concept that they first described in 2012 (ref. 2). The guide represents the latest fruits of Boyden’s partnership with the laboratory of Craig Forest, a mechanical engineer at the Georgia Institute

of Technology in Atlanta who specializes in robotic automation for research.

Most neural recordings involve inserting an electrode in the space between cells to pick up electrical volleys between neurons. Such ‘extracellular recording’ detects outgoing signals but misses the electrical activity inside the cells that determines whether they will fire. This is where whole-cell patch-clamping comes in, a technique that can tap into a neuron’s innards. The delicate procedure “has a very steep learning curve, and even then some people never really get it to work”, says Svoboda.

Whole-cell patch-clamping involves pushing a tiny glass pipette containing a wire ►

► electrode through the brain. In the most common, ‘blind’ version, researchers do this without being able to see the neurons. The scientist must continually apply pressure to push brain matter away from the pipette, but when a rise in electrode resistance indicates that a cell is nearby, they must switch to suction at just the right moment to seal a tiny patch of the neuron’s membrane against the pipette’s super-thin tip. With an additional burst of suction, the researcher can then make a tiny hole in the cell membrane to record the neuron’s activity. Hitting the neuron at the wrong angle, misregulating the pressure and numerous other variables often derail recordings.

“Every step has a certain failure rate, and these multiply throughout the process,” says Boyden. Experienced practitioners report success rates of between 20% and 60%.

Boyden and Forest decided to automate this tricky technique. Their robot does not outperform human experts yet, but its average success rate is around 33% in tests on mice. The device, which runs on the commercial programming platform LabVIEW, only requires researchers to position the animal and the pipette. A computer algorithm then controls the pipette’s internal pressure and its progression through the brain. A company called Neuromatic Devices in Atlanta, Georgia, offers machines based on

Boyden and Forest’s technology, but did not disclose pricing or sales figures to *Nature*.

At the University of Texas at Austin, researchers have created a similar auto-patching system that is controlled in the MATLAB computing environment. This system uses a slightly different algorithm to decide when to start suctioning, and it succeeds in patch-clamping cells in mice about 17% of the time³. Neuroscientist Niraj Desai, who led the team, says that he hopes to incorporate more-sophisticated algorithms.

Some researchers question whether the recording robots will ever surpass the best human experts. “The elements that go into the human’s decisions may be richer than can be captured by the machine,” says neuroscientist Michael Hausser at University College London. But he adds that the technologies could still be a huge boon to novices. Others suggest that the robots could help users of all skill levels in lengthy or complex experiments, in which human fatigue becomes a limiting factor.

At the Allen Institute for Brain Science in Seattle, Washington, researchers have developed an automated system to assist in the even more challenging ‘image-guided’ variant of the technique. In this version, instead of blindly bumping into neurons with a pipette, scientists target specific neurons near the brain’s surface using a two-photon microscope. The procedure requires more coordination than blind

patch-clamping because the scientist must constantly focus the microscope in addition to guiding the pipette and adjusting its internal pressure. “This is a technique that ideally would benefit from having three hands,” says Hausser, an expert in image-guided patching.

The automated system constructs 3D images of the brain region of interest and allows users to digitally select the neuron that they want to record. Then, with the coordinates locked in, the device navigates the pipette into place. For now, researchers still need to patch onto the cell by hand, but Allen Institute neuroscientist and joint team leader Lu Li says that eventually they hope to fully automate the procedure.

Whether these automation systems will be taken up widely by the neuroscience community remains to be seen. Each of the teams has made their code freely available for people to download: Boyden’s group at autopatcher.org; Desai’s team at clm.utexas.edu/robotpatch; and Li’s team at the GitHub repository (go.nature.com/sjgjab). “Our hope is that we can help as many people as possible to answer questions about how neurons compute,” Boyden says. ■

1. Kodandaramaiah, S. B. *et al. Nature Protoc.* **11**, 634–654 (2016).
2. Kodandaramaiah, S. B. *et al. Nature Methods* **9**, 585–587 (2012).
3. Desai, N. S., Siegel, J. J., Taylor, W., Chitwood, R. A. & Johnston, D. J. *Neurophysiol.* **114**, 1331–1345 (2015).

DATA SHARING

Web widget nudges scientists to share their data

Open Data Button launched to encourage public sharing of data sets.

BY DALMEET SINGH CHAWLA

A free, web-based tool that promises to help its users to ask authors of research papers to share their data publicly — and to make such requests publicly trackable — launched in beta version on 7 March. The Open Data Button (opendatabutton.org), a downloadable web-browser extension, can be clicked when a reader is looking at a research paper and wants to see its underlying data, says Joseph McArthur, who is co-leading the project and is assistant director of the policy advocacy group The Right to Research Coalition (R2RC) in London.

When clicked, the button generates a template e-mail that the user can send to the paper’s authors. It asks them to share the data supporting the paper, explains how to do so and — if

the user has typed in the information — states why the data would be useful. All requests are simultaneously posted on the Open Data Button website, where anyone can comment on existing entries to note that they want access to the same data sets.

If the author replies — either with a web link or attached data file — “we ask our users to affirm that this is the data they wanted”, says McArthur. The tool will chase authors once a week for four weeks after a request is filed, he adds, after which entries will be marked as ‘failed’.

The project is mostly funded by a US\$25,000 grant from the non-profit Center for Open Science in Charlottesville, Virginia; the centre has promised to host any data files sent in reply on its own Open Science Framework repository. The aim is to encourage open data sharing,

which is still far from the norm in research, McArthur notes, even though many journals are now asking authors to publish their data alongside their research papers.

“It is good to draw attention to an important problem such as data availability with a one-click gadget,” says Bernd Pulverer, chief editor of *The EMBO Journal* in Heidelberg, Germany. But a button by itself “is probably not quite going to cause the revolution”, he says. There is still a reluctance — especially among biologists — to share data openly, because of fears of being scooped by competition, and because of the extra work required to make data sets open, he notes. Ultimately, McArthur hopes that the button will not be needed as it becomes the norm to share data openly. But progress could be slow — so he thinks that the tool can expect a long and useful life. ■

CAREERS

TOP VALUE Elite researchers identify scientific virtue **p.139**

HINDSIGHT A new PhD graduate counsels her younger self go.nature.com/xqq22d

NATUREJOBS For the latest career listings and advice www.naturejobs.com

MARK HUNT/HUNTSTOCK/CORBIS



Scientists who have disabilities use a growing array of specialized equipment to enable them to carry out research.

DISABILITY AWARENESS

The fight for accessibility

An aisle too narrow, a lab bench too high: the scientific world is a complex place for researchers with disabilities. But many of them find ingenious ways to make it work.

BY ERYN BROWN

Henry “Hoby” Wedler caught the chemistry bug as a high-school student in Sonoma County, California. He loved to think about atoms and how they fit together to make molecules. Although he wanted to enrol in his school’s advanced chemistry course, he faced a frustrating problem: he has been completely blind since birth, and his teacher thought that he might find it too hard. The field was too visual, she told him.

Nonsense, Wedler thought. “I said, ‘No one

can see atoms; it’s completely cerebral.” Now 28, he is studying for a PhD in organic chemistry at the University of California, Davis, and is considering teaching offers. His once-doubting school instructor has become one of his greatest allies.

It’s not an easy road, but early-career scientists who face physical challenges such as blindness, deafness or paralysis are building varied and rewarding careers. They work in academic, government and industrial research as teachers, consultants and other occupations.

Success requires desire, grit and ingenuity.

Researchers who have trouble seeing, hearing or with mobility need creative workarounds in the lab and field. They may have to design their own equipment. They need a crew of friends, peers and mentors who can provide support. And they must seek work that capitalizes on their strengths, accepting that some assignments may be beyond their reach — at least for now.

“You have to really want to be a scientist a lot, or it won’t happen,” says Richard Mankin, a research entomologist for the US Department of Agriculture (USDA) in Gainesville, Florida, who wears braces on his legs and ►

► relies on crutches to walk.

Despite a patchwork quilt of policies and guidelines that are meant to broaden employment opportunities, disabled people still have trouble finding work. A 2014 analysis by the Campaign for Science and Engineering, which promotes science in the United Kingdom, reported¹ that fewer than half of the nation's 5.2 million people of working age with disabilities had jobs between 2010 and 2011. They were more than twice as likely as peers without disabilities to report working part-time, and about half as likely to have jobs in science, technology, engineering and mathematics.

Similarly, in 2015, the US National Science Foundation reported² that about one in nine scientists aged 75 or younger in the United States had a disability. They, too, were more than twice as likely to be out of the labour force than their peers without disabilities.

As Wedler found, accessibility issues crop up long before a scientist enters the workforce. German-born climatologist Imke Durre, blind since early childhood, says that her father, a computer scientist, created a Braille word-processing program in the early 1980s to help her to do school work. Such technology is widespread today, says Durre, who works for the US National Oceanic and Atmospheric Administration.

But she says that there are still practical hurdles to using assistive tools, particularly for students. Even when Braille versions of textbooks are available, teachers may not know about them or be able to order them in time for a class. Figures, tables and graphs typically aren't translated into Braille, so a student with a visual impairment often needs to collaborate with a sighted colleague to interpret visual data — a process that may not go smoothly, says Durre. On occasion, she says with a laugh, she has shown up to a meeting carrying blank pages instead of the hard copies of graphs and tables that she thought she had prepared.

A FLEXIBLE SPACE

Mundane tasks pose barriers, too. James McNutt, who uses a wheelchair and studies the history of science at Queen's University in Kingston, Canada, set out last year to record some of the difficulties that he faces as he travels around the university (see go.nature.com/oaablb). With a video camera attached to his wheelchair, he attempted to open doors, operate lifts and visit washrooms. Buttons, switches and knobs were often out of reach. Doors wouldn't open. Passageways were too narrow or too winding to accommodate a wheelchair easily.

"Quite often, the planning people don't have an idea of what is and is not accessible," says McNutt, who has cerebral palsy. "They don't know how big the wheelchair is. They don't have any idea what it's like."

Audrey Kobayashi, a geographer at Queen's University and a member of the school's committee on campus accessibility, says that the situation has improved greatly over the past two

- The STEM Disability Advisory Committee, a partnership between the Royal Society and other scientific groups in the United Kingdom, pushes to improve funding for scientists with disabilities, and hosted a conference in March focusing on the transition from study to the workplace.
- The American Association for the Advancement of Science led early on in making conferences accessible, and sponsors Entry Point!, a programme that places young scientists with disabilities at internships with university, government and industry labs.
- Sang-Mook Lee's Quality of Life Technology initiative at Seoul National University, funded

decades. There are now clear lines of responsibility to ensure that students with disabilities get any help they need, and the movement is garnering attention (see 'A worldwide wave of awareness'). But Kobayashi, who uses a wheelchair at work because of a neurological disorder called transverse myelitis, says that a lot of challenges remain. "We're trying to make a barrier-free campus, but it's slow," she says.

The difficulties don't end after receiving a degree. Researchers who have a disability and want to work in a scientific field must first ensure that physical adjustments are made to labs and other workplaces to facilitate access. These can include redesigning lab sinks to accommodate a wheelchair, posting emergency instructions in Braille and checking that doorways to halls with lifts or ramps don't lock automatically and block exits.

In the United States, the Americans with Disabilities Act, enacted in 1990, requires accommodations in public spaces. The UK Disability Discrimination Act (1995) and Canada's Accessibility for Ontarians with Disabilities Act (2005) impose similar requirements for their regions. But the laws often do not require retrofitting, and thus many labs that were built before the legislation took effect remain difficult or impossible to navigate for scientists who have mobility issues: aisles are too narrow, tables are too tall and eyewash stations are tucked into inconvenient corners.

Some scientists are attacking the problems on their own. Neuroscientist Bradley Duerstock, who has been quadriplegic for more than 25 years, developed a disability-friendly lab space at Purdue University's Center for Paralysis Research in West Lafayette, Indiana, where he researches assistive technologies. With grants

GROUP ACCOMMODATION

A worldwide wave of awareness

by the South Korean government, promotes adaptive technologies for disabled students.

- Analytical chemist Karl Booksh at the University of Delaware in Newark runs a programme for disabled undergraduates that prepares them for graduate work.
- The American Chemical Society's Chemists With Disabilities committee has published a guide to teaching disabled chemistry high-school, college and graduate students³.
- The International Association for Geoscience Diversity, in Cincinnati, Ohio, seeks to improve accessibility for students and scientists with disabilities in the Earth sciences. **E.B.**

from the US National Institutes of Health and other sources, he adapted the 'kitchen work triangle' — a home-design layout concept that imagines the stove, sink and refrigerator as the shape's corners — into a 'wet-laboratory work triangle' defined by the lab sink, lab bench and fume hood. Years ago, he designed a microscope that lets the user control illumination, focus and exposure through a computer interface, rather than tricky-to-operate knobs (now, such microscopes are commercially available).

Scientists with disabilities are also installing light switches that are positioned at an accessible height and that are labelled for people who rely on Braille, and they use adjustable-height lab benches and other accommodations.

Back in California, Wedler gives credit to his adviser, chemist Dean Tantillo, for hiring him as an undergraduate and for making the lab an easier place to work. With the help of US\$30,000 in supplemental funding attached to his US National Science Foundation graduate fellowship, Wedler and other members of Tantillo's team developed a 3D-printing solution that produces tactile models of molecular structures. Different shapes and textures in the models represent different atoms and bonds, and Braille notations describe bond angles and bond lengths. It takes a couple of days to generate each printed model, but Wedler doesn't mind: the homegrown technology allows him to 'feel' the outcomes of his calculations so that he can verify his work on his own.

When Duerstock and Susan Mendrysa, then a colleague at Purdue's Institute for Accessible Science, polled coworkers to learn what qualities make a lab most accommodating, they found that working with an established, well-funded principal investigator was key: such scientists were more likely to give staff with disabilities the extra time they might need to complete their studies or to publish papers.

Colleagues can also help on a smaller scale. Born with muscles missing in his legs and

"Quite often, the planning people don't have an idea of what is and is not accessible."



Entomologist Richard Mankin (right) does work for the US Department of Agriculture in Guam.

AUBREY MOORE

arms, Mankin has conducted field research internationally since the 1970s, walking with crutches or crawling along the ground to study the sounds and vibrations that insects make in various locations. He seldom works in the field alone, and he keeps his trips short. He asks those who accompany him to manage tasks that he cannot perform, such as carrying equipment and climbing trees.

Physical barriers are not the only obstacles: bias can also be an issue. Jae-Hyeon Parq, a postdoctoral researcher at Seoul National University, who has used a wheelchair since sustaining a spinal injury as an undergraduate, worries that his disability will make it hard for him to find a job. Trained as a physicist, Parq now works in the lab of marine geologist Sang-Mook Lee, who has been trying to improve conditions for scientists with disabilities since 2006, when he was paralysed in a car accident.

“Most people, especially in Korea, don’t understand the diversity of disabled people,” Parq says. “They judge what I can and what I can’t do from my appearance.” If Parq can’t get a permanent job, he says, he will continue to work for Lee.

PUBLIC SECRETS

Those whose disabilities aren’t as immediately obvious face a different, yet related problem: whether to tell potential employers. “One of the most common questions I get is, should I say on my CV that I’m deaf?” says biochemist Annemarie Ross of the Rochester Institute of Technology’s National Technical Institute for the Deaf in New York. Ross, who is hearing-impaired, tells students that it is their choice — there is no clear advantage for applicants who do or don’t reveal a disability.

But it’s a challenge that must be resolved, she says. “A big barrier in general for our students

are the employers. They think, ‘If a worker can’t hear a fire alarm, how do we make sure they’re safe? If they stay behind in a burning lab, we could be liable.’” Often, job candidates must persuade employers to reframe their assumptions in interviews, Ross says. Those with hearing disorders, for example, can see the strobe lights on many modern fire-alarm systems. By the same token, scientists in a lab don’t spend much time doing physical tasks.

“I was always having to persuade people I could do things from a wheelchair,” says Karl Booksh, an analytical chemist at the University of Delaware in Newark who experienced a spinal cord injury in university. “The way I convinced most of them was pointing out that the most successful faculty members didn’t know where the pipettes were to begin with — that the key to success was writing papers and proposals.”

Some scientists with disabilities have reframed their impairment as a positive attribute: they say that coping with the challenges of everyday life has helped them to develop unusual skills and expertise. Wedler, for instance, says that navigating town trained his brain to make spot-on mental maps. A similar sort of spatial thinking helps him with organic chemistry. “I was thinking in terms of feet and miles, but there’s no reason you can’t shrink that down to ångströms,” he says. “In terms of doing the problems, I might have an advantage over my sighted peers.”

Mankin is dubious that the stigma against those with disabilities will ever fade completely. He is president of the Foundation for Science and Disability, which sponsors a grant programme that supports the research of graduate students with disabilities. But, he says, he doesn’t think of himself as disabled.

He is an enthusiast whose voice crackles with excitement when he talks about his work. He is studying psyllids, insects that cause a tree-damaging disease that threatens Florida’s \$10-billion citrus industry, and he has been developing systems that use vibrations to lure and trap male psyllids to prevent them from mating with females nearby. The approach could offer an alternative to pesticides, and has attracted the attention of federal legislators.

“Being a scientist has been lots of fun,” Mankin says. “I’ve done things that I hope have benefitted humanity. This is what I always wanted to do.” ■

Eryn Brown is a freelance writer in Los Angeles, California.

1. *Improving diversity in STEM* (Campaign for Science and Engineering & King’s College London, 2014).
2. *Women, minorities, and persons with disabilities in science and engineering* (National Science Foundation, 2015).
3. Miner, D. L., Nieman, R., Swanson, A. B. & Woods, M. (eds) *Teaching Chemistry to Students with Disabilities* 4th edn (American Chemical Society, 2001).

CHARACTER TRAITS

Scientific virtue

Honesty and curiosity are the most important traits underlying excellent science, according to a survey of around 400 members of elite US scientific societies, such as the National Academy of Sciences. A pilot study led by survey co-organizer Robert Pennock, a philosopher at Michigan State University in East Lansing, had previously identified the ten most widely held values among scientists who have been honoured by their peers for being exemplary. Although honesty and curiosity dominated, these virtues also included perseverance, objectivity and the willingness to abandon a preferred hypothesis in the face of conflicting evidence (see ‘Core values’).

Little empirical research has been done to learn what traits scientists value most in one another, says Pennock, and this work indicates a high level of consensus among elite US researchers about what is important for the practice of science. He thinks that training programmes that emphasize such shared scientific values are likely to be more effective than are those that focus on compliance with official rules of behaviour; 94% of the scientists surveyed felt that scientific virtues can be learned.

About four in five of those surveyed feel that today’s trainees share the scientific values that they themselves held when training, and 88% take candidates’ scientific character traits into account when recruiting lab members. The team members presented their preliminary results at a meeting of the American Association for the Advancement of Science in February (see go.nature.com/o4urjl), and they plan to publish full results from a sampling of 500 established scientists, in addition to a similar-sized group of early-career scientists, in upcoming months.

CORE VALUES

Elite scientists were asked which three values they consider to be the most important.



*Willingness to abandon a preferred hypothesis when faced with conflicting results.

CHOICES, IN SEQUENTIAL ORDER

Into the unknown.

BY KARLO YEAGER RODRÍGUEZ

SINGLE-AXIS DESIGNS:

THE KEY TO LIFE! (COPYRIGHT 2349).
START

PLANET: EARTH / TAU CETI E
Tau Ceti e.

REDIRECTING TO XENOBIOTA...

You're killing me, down in this tunnel you've made, but I feel like I can only talk to you. Strange, huh? I feel euphoric. I wonder if it's the low oxygen levels on the planet's surface. I feel it's because of you. I'm 12 light years away from home, paralysed in a hole in the ground, on a hell-planet, but you're everything I ever dreamed in my life.

CAN IT MOVE? PRESENT / NOT PRESENT
Present.

SYMMETRY: RADIAL OR PENTAMERISM /
BILATERAL
Bilateral.

Because of you, the promise of you, Vee and Mom agreed. The one time they agreed on something, y'know? They joined forces, but I was ready. I told them Pop would've jumped at the chance I was given.

Xenobiologist! An advance team. Pop would have been so proud! The only space exploration he knew was contained in the paperbacks he read until he couldn't read any longer. When that happened, I read to him. The breathing machines shushed me, the rattle and hiss of air pumped in and out of Pop's lungs. His eyes closed, I'm sure he dreamed of the vast frontier. And one day — one day, the machines stuttered into stillness. I knew the time had come. With Pop gone, why did I have to stay Earthbound?

After I beam the dichotomous key data back, I'll suggest naming you after my Pop. *Xenopendra Pérezii* has a good ring.

When I mentioned his wishes, Vee and Mom said it wasn't fair. What does fair have to do with anything?

PANARTHROPODA

JOINTED LEGS? YES / NO

Yes.

HOW MANY ANTENNAE? TWO PAIRS /
ONE PAIR OR NONE

It only has one pair.

It's getting dark in here. How's the saying go? Better to light a viewscreen than to curse the darkness, huh? Good thing my suit's helmet has a heads-up display. Best thing, the display — I have something to do while you're gone. Auxiliary team needs to know about you.

Thank you for pinning me down. Probably break my neck. I'm dead either way.



My cheek itches.

Where are —? Oh — you're fast! You must not need visible light to see, huh?

Oh, I — I guess maybe I moved, but that glued me to the spot. Won't move anymore.

I can feel them wriggle around, now. I wonder if they'll tickle when they hatch?

NUMBER OF LEGS? THREE PAIRS /
FOUR OR MORE PAIRS

Four or more pairs.

NUMBER OF LEGS PER SEGMENT? TWO
PAIRS / ONE PAIR

One pair.

Can't feel my fingers. There's a hot breeze across my belly, though. The crosswind tells me you must have dug out another exit. The airflow leaves my mouth full of sand. I can feel it across my belly, too — suit's torn open.

The diagnostic module doesn't respond. You must have damaged it when you ripped it open. It could've analysed for poison. Vitals circuit's running. Redundancies: good. The beacon's activated if vitals drop to null.

Fangs — forcipules, yes, not fangs — must deliver a weak toxin. Is the euphoria caused by it alone, or is it a factor? What about low oxygen? Temperature?

I must've pissed myself — I can smell it.

Pressure like a clenching fist on my abdomen. Now, it unclenches and flutters open, moves, clenches again.

Why am I so — accepting? I know — my brain knows — your young will devour me. Why can't I feel panic? Fear? I should stop — why can't I stop talking to you?

NUMBER OF TRUNK SEGMENTS? 12 /
MORE THAN 12

More than 12. Much more.

FORCIPULES? YES / NO

Yes.

Vee never listened — 'no' meant getting what she wanted suffered a setback. She pushed and pushed and pushed — always what she wanted. Exhausting. She wanted to

have kids — she tells me this the night before. Incredible, right? I'm ready to go to the facility, to get scrubbed, suited up, loaded into a high-tech version of a tin can and shot across the vast reaches of space. She wants to talk babies.

Strange — before, Vee had shouted down to me from her place on the corporate ladder: a few more years! I ignored her.

Where are you? I hear a scraping sound at the far end of the tunnel — are you digging? Are you going to leave me here? I suppose I should expect it. I would have thought you might watch over your babies.

No. Larvae — the right word is larvae. Certain species coil around their egg cluster. Not you.

So — babies. Oh, you should have seen Vee break stuff! She even threatened me at knifepoint, blocked my way out of the house. The security team made sure I got to the facility in time, though. They hustled me out of the house, through the prep stations. It's funny. Now I remember: they wanted to make sure I didn't bring any bugs with me. Hilarious, right?

CHILOPODA? YES / NO

No.

RECALCULATING...

RECALCULATING...

RECALCULATING...

SIZE: A METRE OR LESS / MORE THAN
A METRE

More than a metre.

CURVED TAIL, WITH OVIPOSITOR?
YES / NO

That an ovipositor in your pocket? Or are you happy to see me?

**'THETA OVIPOSITOR AIR POCKET
ARE YOU HAPPY TWO SEAMY' IS NOT
ACCEPTED. PLEASE ANSWER IN 'YES'
OR 'NO' FORMAT**

Yes.

Yes.

It's happening.

They do tickle and my breath is ragged as I suppress laughter. The viewscreen flashes once, flickers and goes out, switches to auxiliary power. I am part of something bigger, something more, now.

**ALERT! DANGEROUS XENIFORM
ACTIVATING RESCUE BEACON
SOS SOS SOS ■**

Karlo Yeager Rodríguez is from Puerto Rico, but moved to Baltimore, Maryland, a few years back. His fiction has appeared in PULP Literature and Clowns: The Unlikely Coulrophobia Remix.

ILLUSTRATION BY JACEY